COOPERATION IN RESEARCH

STAFF MEMBERS
RESEARCH ASSOCIATES

In Memory of
Remington Kellogg

1892 1969

Mammalogist
Paleontologist

John C. Merriam

Dec. 14, 1938

COOPERATION IN RESEARCH

HARRIS & EWING

John C Merriam

March. 1932

Q
111
C28
1938
VPAL

COOPERATION IN RESEARCH

BY

STAFF MEMBERS AND RESEARCH ASSOCIATES
THE CARNEGIE INSTITUTION OF WASHINGTON

GIFT
From the Library of
REMINGTON KELLOGG

Vertebrate Paleontology
U. S. National Museum

Published by The Carnegie Institution of Washington
Washington, D. C.
1938

SMITHSONIAN
MAR 17 1993
LIBRARIES

CARNEGIE INSTITUTION OF WASHINGTON PUBLICATION NO. 501

THE WAVERLY PRESS, BALTIMORE, MD.
THE STANDARD ENGRAVING COMPANY, WASHINGTON, D. C.

FOREWORD

President John Campbell Merriam will retire, at his own request, on December 31 in order that he may be free to devote his time to research and to writing. The Board of Trustees of the Carnegie Institution of Washington, on December 10, 1937, appointed him President Emeritus, effective January 1, 1939, with provision for support of his studies.

Shortly after this action a small group of the members of the Institution suggested that a dinner be given in his honor and that a testimonial volume of papers written by members and research associates of the Institution be presented to him at that time in appreciation of his farsighted leadership. The Trustees approved the proposal and appointed a small committee to cooperate in the project. After thorough discussion, the committee concluded that the volume should comprise papers on current problems, conducted either entirely or partly within the Institution, and illustrative of the principles followed by Dr. Merriam as President during the past eighteen years. It was agreed that these papers might serve, in part, to indicate the effectiveness of cooperation in investigation and, at the same time, epitomize the nature of the activities of the Institution in various branches of science.

Throughout his administration Dr. Merriam has envisaged the Institution as a unified organization devoted to the conduct of research in science and to the interpretation of the results of research in terms of human values. He has stressed the need for cooperative attack on problems too large to be solved by a single observer and has initiated many projects in which teamwork by different groups has been essential to success. This extension of the activities of the Institution to new fields of science, peripheral to those of its departments, has proved extremely fruitful. Convinced that the public is entitled to know of the work of the Institution, he has sought to furnish this information through annual exhibitions of current work, through public lectures, and through preparation of news releases, news bulletins, magazine articles, and books on special items of general interest arising from the research activities of the Institution.

Each author, in writing his article for this volume, has endeavored to record the results obtained by himself or his group. Each paper serves to focus the attention of the reader on a specific problem, so presented that the necessary background is sketched to render its relation to the general field apparent and thus, by way of illustration, to encompass the entire field. These papers alone demonstrate, better than pages of argument, the value of cooperative endeavor in science.

It is hoped that this volume will express to President Merriam the appreciation felt by the members and research associates of the Institution for his continued inspiring guidance of their work and for his stimulating personal interest in their problems.

Committee from the Board of Trustees:
Robert Woods Bliss
Lewis H. Weed
Frederic A. Delano, *Chairman*

Committee from the Staff:
L. H. Adams
Walter S. Adams
Frank F. Bunker
J. A. Fleming
W. M. Gilbert
A. V. Kidder
H. A. Spoehr
George L. Streeter
F. E. Wright, *Chairman*

December 8, 1938

CONTENTS

PHYSICAL SCIENCES

An Adventure in Scientific Collaboration 3
Arthur L. Day
The Significance of Pressure and of Volume in Geophysical Investigations 37
L. H. Adams
Water in Geological Processes 49
George W. Morey
The Surface of the Moon 59
F. E. Wright
The Electric Eye in Astronomy 75
Joël Stebbins
Explorations in the Realm of the Nebulæ 91
Edwin Hubble
Solar and Terrestrial Relationships 103
Seth B. Nicholson
Knowledge of the Planets in 1938 115
Theodore Dunham, Jr.
Sunspots and Stellar Distances 135
Walter S. Adams
The Brightness of Stars: Standards of Magnitude and Color .. 149
Frederick H. Seares
Magnetism and the Atomic Nucleus 161
M. A. Tuve, L. R. Hafstad, and N. P. Heydenburg
Universal Aspects of Atmospheric Electricity 183
O. H. Gish
The General Magnetic Field of the Earth and Its Secular Variation 205
J. A. Fleming
The Ephemeral Variations of the Earth's Magnetism 223
L. V. Berkner and A. G. McNish

BIOLOGICAL SCIENCES

The Cooperative Researches of the Nutrition Laboratory 251
Francis G. Benedict

THE CHANGING ORGANISM.. 259
Oscar Riddle
OBSERVATIONS ON EVOLUTIONARY CHANGES IN THE CHROMOSOMES OF SCIARA (DIPTERA).. 275
C. W. Metz
EIGHTEEN YEARS OF RESEARCH ON THE GENE..................... 295
M. Demerec
FIFTEEN-YEAR BREEDING RECORDS OF $2n + 1$ TYPES IN DATURA STRAMONIUM 315
Albert F. Blakeslee and Amos G. Avery
WHAT AND WHY IS LEUKEMIA?—A MEDICO-BIOLOGICAL STUDY ON MICE.. 353
E. C. MacDowell
STUDIES ON WHITE BLOOD CELLS................................ 369
Warren H. Lewis and Margaret R. Lewis
EARLY DEVELOPMENT OF THE PRIMITIVE MESOBLAST IN EMBRYOS OF THE RHESUS MONKEY.. 383
Chester H. Heuser
ATTEMPTS TO INDUCE OVULATION IN THE ADULT RHESUS MONKEY... 389
Carl G. Hartman
CHARACTERISTICS OF THE PRIMATE EGG IMMEDIATELY PRECEDING ITS ATTACHMENT TO THE UTERINE WALL...................... 397
George L. Streeter
THE INFLUENCE OF NUTRITION UPON THE CHEMICAL COMPOSITION OF THE NORMAL BODY.. 415
Henry C. Sherman
THE EFFECTS OF VOLUNTARY CHANGES IN BREATHING ON HUMAN RESPIRATORY EXCHANGE.................................... 425
Thorne M. Carpenter
FORM, FORCES, AND FUNCTION IN PLANTS....................... 435
H. A. Spoehr

HISTORY

THE SCIENTIFIC BASIS OF THE HISTORY OF SCIENCE................ 465
George Sarton
THE SUBSTANCE OF AMERICAN HISTORY.......................... 483
Waldo G. Leland
THE CORRELATION OF MAYA AND CHRISTIAN CHRONOLOGIES........ 493
A. V. Kidder and J. E. Thompson
RACE AND CLASS IN YUCATAN.................................... 511
Robert Redfield

The Maya New Empire 533
Sylvanus G. Morley
The Maya Indians of Yucatan 567
Morris Steggerda
Fray Diego de Landa and the Problem of Idolatry in Yucatan .. 585
France V. Scholes and Ralph L. Roys

PALÆONTOLOGY AND GEOLOGY

Studies Relating to Early Man in America 623
E. B. Howard
Ancient Forests of Oregon: A Study of Earth History in Western America 631
Ralph W. Chaney
Adaptation of Structure to Function in Whales 649
Remington Kellogg
Structures in Modern Sediments Aid in Interpreting Ancient Rocks 683
Edwin D. McKee
Earth History of a Portion of the Pacific Northwest 695
J. P. Buwalda

PHILOSOPHY AND INTERPRETATION

Cooperative Research, Its Conduct and Interpretation 713
Frank F. Bunker
Transmuting Science into Conservation 753
Newton Bishop Drury
John Campbell Merriam as Scientist and Philosopher 765
Chester Stock

PHYSICAL SCIENCES

AN ADVENTURE IN SCIENTIFIC COLLABORATION

ARTHUR L. DAY

Chairman, Advisory Committee in Seismology
Formerly Director, Geophysical Laboratory
The Carnegie Institution of Washington

Among the policies which have distinguished the present administration of the Carnegie Institution of Washington, cooperation with other productive research agencies is outstanding. President Merriam may have had a deliberate purpose to broaden the base of American research activity by interesting more agencies in some of the larger research problems, or perhaps he was merely recognizing the fact that certain pressing problems, too large or too diversified even for the Carnegie Institution to undertake alone, do not require to be laid aside for that reason, but can and should be attacked by an organization commensurate with their scope. A forward-looking research spirit does not recognize such limitations as necessarily prohibitive. It is significant that Carnegie Institution was originally dedicated to the pursuit of the larger research problems which, because of their very scope, were less likely to be undertaken by any other agency. Perhaps there need be no surprise that among these larger problems some would be found requiring the combined effort of several cooperating agencies for their successful prosecution and that this fact would prove no bar to their favorable consideration.

Without attempting to assign reasons for it, it is certainly true that the status of American seismology, about the year 1920, was uninspiring, and that this fact attracted the attention and stimulated the initiative of President Merriam to try to "do something about it." The problem was not entirely new to the Institution, for one of the most thorough investigations ever undertaken upon any single earthquake, namely the study of the great San Francisco earthquake of 1906, was sponsored by the Carnegie Institution and the four splendid volumes containing the results of that investigation still stand as a monument to that early effort. Nevertheless, such seismologic studies were not continued after the completion of these volumes and in 1920 none of the larger research agencies in the United States were devoting major efforts to this field of activity. Indeed it was commonly said at that time that no seismologic station existed in the United States which was entitled to an A rating internationally. Such stations as were in existence then were maintained with the purpose of recording and reporting distant earthquakes of public interest or for the instruction of students of seismology in the universities.

This fact is the more noteworthy in view of the local situation then prevailing on the west coast of the United States, which had but lately passed through the catastrophic earthquake of 1906 which had caused great destruction in the business center of San Francisco, and where earthquakes of lesser intensity continued to occur from time to time along the coast from Point Conception to the Mexican boundary. These local ground movements, which might again be expected to culminate in widespread destruction whenever they happened to occur in or near centers of population, were not being studied by any agency then existing. Indeed such study was not encouraged by the local inhabitants who might be expected to be most interested, because of the view widely entertained on the west coast that property interests might be adversely affected if public attention were drawn to the earthquake hazard.

It was with this background that consideration came to be given in 1920 to the serious study of earth movements within the United States.

It is appropriate to record the fact that Carnegie Institution was not the first to give serious consideration to the situation on the west coast, but it was, I think, the first to make a serious effort to meet it. The departments of geology both at Berkeley and at Stanford University were thoroughly cognizant of these threatening conditions and had even brought them to public attention through occasional publications in the scientific journals. Also in 1920 the Seismological Society of America, a Pacific Coast organization of those interested in this subject, had been in existence for a decade, and had encouraged research in this field of activity within the narrow bounds of local academic and engineering interest in the problem. It also happened that Mr. H. O. Wood had lately prepared and printed[1] a definite plan for the investigation of west coast earthquakes indicating the number and appropriate location of stations necessary for a thorough survey of California earth movements and of the instruments necessary for such a survey.

Wood's plan was brought to the attention of various research agencies at that time in the hope that one or more of them might be interested to take up the investigation and to support it with adequate resources of men and money, but whether the agencies approached happened to have their centers of activity in the east, while this was presented rather definitely as a west coast problem, or whether the very large expenditure, something over a million dollars, suggested for the competent equipment of such an undertaking served to frighten away some who might otherwise have been interested, is not of particular importance now. The important fact is that it came into the hands of President Merriam eventually, and that he was sufficiently interested to have a study made of the proposed undertaking with reference to its scope and practicability, and to this end he appointed in 1921 an Advisory Committee in Seismology, of which the writer was asked to serve as chairman, and a number of distinguished scientists,[2] known to possess an active interest in studying the stability of the earth's crust, were associated with him. Although the majority of this Committee consisted of west coast men with a direct interest in the local problem and first hand information on its history and outlook, the survey was made with a broad purpose of increasing our knowledge of earth movements within the United States.

After a year of observation and inquiry this Committee reported in October 1921 in favor of recommending to the Carnegie Institution of Washington that it take the initiative in a systematic study of regions of instability within the United States beginning with the west coast. Specifically its major conclusions took the following form:

---

[1] The Earthquake Problem in the Western United States, Bull. Seismol. Soc. Amer., December 1916.

[2] The Advisory Committee in Seismology was made up as follows: Dr. J. A. Anderson, astrophysicist at Mount Wilson Observatory; Dr. Ralph Arnold, oil-well engineer, formerly of the United States Geological Survey; Dr. W. W. Campbell, then president of the University of California; Dr. Arthur L. Day (chairman), director of the Geophysical Laboratory of the Carnegie Institution of Washington; Professor A. C. Lawson, professor of geology in the University of California; Professor R. A. Millikan, chairman of the Executive Committee of California Institute of Technology; Professor Harry Fielding Reid, professor of geology in Johns Hopkins University; Professor Bailey Willis, professor of geology in Leland Stanford University.

> The proposal of the Carnegie Institution of Washington to enter the field of seismology has the approval of the Advisory Committee in Seismology for several important and, in the opinion of the committee, opportune reasons.
>
> (1) This country has not hitherto taken an active part in seismologic research when compared, for example, with England, Germany, or Japan.
>
> (2) In the State of California there is probably a more favorable opportunity for the study of crustal movements in great variety than in any other region, save possibly Japan. Furthermore, if the landslips there are due in whole or in part to accumulating stresses caused by crustal drift or otherwise, then the present is is a more opportune time for their study than the period immediately following a release of these stresses (earthquake).
>
> (3) As research is organized in the United States, a reasonably comprehensive study of California earth movements would require the participation of several agencies which would probably not undertake the task independently, but which have indicated a desire to cooperate with the Carnegie Institution in such a study. It is of interest to note that, since the action of the Institution became known, resolutions of indorsement and support have been passed by the San Francisco section of the American Institute of Mining Engineers (May 17), by the Board of Directors of the Seismological Society of America (May 25), by the Executive Committee of the American Geophysical Union (September 17), and by the Commonwealth Club of San Francisco (November 17).
>
> (4) Studies of local movements as opposed to teleseismic observations have not been systematically undertaken hitherto (except to some extent in Japan). They appear to offer a fertile field of research and to promise information of considerable economic value, particularly to the West Coast region of the United States.

It is pertinent to note that paragraph 3 of the conclusions of the Committee specifically indicated the need for an intimately collaborating organization of a kind then somewhat novel in American research. The scope of this collaboration and the particular research agencies invited to participate in it are set forth in Year Book No. 21 (page 390) of the Carnegie Institution as follows:

> In particular it was recommended that the Ukiah and Lick Observatories be invited to continue and extend their observations of latitude for the purpose of establishing (or disproving) a northward crustal creep or drift which had been indicated by earlier observations; that the U. S. Coast and Geodetic Survey be invited to extend its system of primary triangulation and precise levels until no considerable area within the various zones of movement in California can suffer displacement without the possibility of establishing its direction and magnitude; that the U. S. Geological Survey, in collaboration with the California universities

and geological societies, be invited to organize geologic studies of the regions in which the more active faults occur; that the California Institute of Technology, the Mount Wilson Observatory, the U. S. Bureau of Standards (Washington), and any others interested, be invited to aid in the development of instruments more suitable than any now in use for recording and analyzing local slips and tremors; and finally, that the effort be made to obtain from the Navy Department deep-sea soundings off the west coast of California to establish the precise location of the continental shelf and any other conspicuous fault scarps adjacent to the land areas in which active faults are found.

These recommendations were considered favorably by the Trustees of the Institution and the Advisory Committee was asked to continue its organization for another year for the purpose of arranging with the institutions mentioned a practicable basis of cooperation through which the various purposes designated above might best be accomplished. The task was accepted by the Committee and during the following year the effort was made to organize working arrangements in accord with the plans above outlined. It was significant, and extremely encouraging to the Advisory Committee, that without a single exception every institution invited was glad to join in a systematic study of this kind and ready to aid to the fullest extent of its power. The organization—the first, be it said, of this magnitude which had been attempted in American research—promised to be a most effective one. This is a tribute not only to the scientific soundness of the problem under consideration, but to the readiness of American scientific bureaus and institutions to advance research even when not in the direct line of their immediate program.

The detailed plans for the stations and instrumental equipment which might prove necessary for the study of ground movements in southern California were intrusted to Mr. H. O. Wood, who had prepared the original plan for California studies in 1916 and who had lately been appointed research associate of Carnegie Institution for this purpose. Wood brought to this undertaking considerable actual field experience obtained during a residence of seven or eight years on the island of Hawaii, where he had inaugurated a systematic study of the local earthquakes emanating from the volcanoes of that island. As a study of local earthquakes this experience was unique in American seismology at that time, and through Wood's participation the present enterprise gained the full advantage of that experience. It also happened that Wood was a geologist by training and that he had had an active share in the investigation of the San Francisco earthquake of 1906, and so was in every way peculiarly fitted to take a leading part in the present undertaking.

Perhaps it is appropriate to say here that the Advisory Committee in Seismology has been continued from year to year since 1921 without change of personnel except for the addition in 1926 of Dr. J. P. Buwalda, head of the Department of Geology at California Institute of Technology, and the death recently of Dr. W. W. Campbell. Mr. Wood's appointment as research associate in seismology also has continued down to the present time (1938). The continuity of the program as well as the intensive study of appropriate means for carrying it out in both laboratory and field are in no small degree due to the experience and enthusiastic cooperation of Mr. Wood.

The undertaking was actually launched in several directions simultaneously in the year 1922. A subcommittee, under the chairmanship of Professor Bailey Willis of this Committee, with whom were associated Professor Lawson and Dr. Arnold, also members of this Committee, Mr. F. P. Vickery of the then Southern Branch of the University of California, and Dr. L. S. Noble of the United States Geological Survey, undertook the task of preparing a plan for a thorough geological study of the faulted regions of California. Dr. Willis himself undertook the responsibility for the northern half of the state and Mr. Wood and Dr. Noble for the less studied southern portion. There resulted from these studies a complete fault map of the state of California upon which were indicated all the faults known or suspected up to that time. This map was completed in 1923 and was published by the Seismological Society of America as a supplement to its quarterly journal in March of that year. As the first of its kind the map has been found to contain some inaccuracies of fault location and of classification, but these are entirely insignificant in comparison with its general usefulness for seismological study and geological interpretation since that time.

Simultaneously with these geological studies it was deemed indispensable that essential landmarks throughout the faulted area in the west coast region should be located with the highest precision to serve as points of reference or indices of movement for the displacements which might subsequently occur, and which were to form the major subject of study of this group. For this purpose the cooperation of the United States Coast and Geodetic Survey, the responsible government agency for primary triangulation in the United States, was solicited and its assistance promised. The active interest of representatives of the state of California in Congress brought about an appropriation of $15,000 for the fiscal year beginning July 1, 1922, with which to inaugurate this work. The plan contemplated the immediate survey of a system of primary triangles determined with the greatest possible precision extending from San Francisco Bay across the important zone of movement of 1906 to the older formations beyond the Sierra at Reno, Nevada. Following this, if further appropriations could be made available, it was proposed to continue the primary triangulation southward to near the Mexican boundary and thence again to the eastward across the faulted zone of southern California to more stable terrane in the Colorado River basin. This program required three years but eventually was carried out as planned with the continued support of the government through the Coast and Geodetic Survey.

While these plans were making to establish a network of monuments in the west coast area, through which the direction and magnitude of any subsequent ground movement could be established with certainty, it was thought to be equally necessary to ascertain the direction and extent of any continuation of the fault zones beneath the immediately adjacent ocean area. Accordingly the aid of the Hydrographic Office of the Navy Department was solicited in the prosecution of our program. It is possible that the interest of the Navy Department was somewhat enhanced by the fact that during the late stages of the Great War an echo-sounding device had been developed which permitted soundings to be made at short intervals from a moving vessel. It had therefore become practicable for the first time to make a contour map of the ocean floor within reasonable limits of time and expense.

In pursuit of this plan the Navy equipped two destroyers with the echo-sounding device and in October 1922 began the series of 5000 new soundings extending outward from the coast line to the continental shelf and along the coast from San Francisco Bay to the Mexican boundary, in all an area of the sea floor amounting to some 34,000 square miles. The soundings were connected up into contour lines by the Geological Survey and the Hydrographic Office and eventually published by the Navy Department as the first successful contour map of an ocean floor area.

Later these same contours were added to the fault map of the adjacent mainland of California, already referred to, and so yielded a reasonably complete record of the surface configuration of this west coast area from the Sierra Nevada outward to the continental shelf.

Considered as a problem in crustal loading it is interesting to note that the difference in level from the summit of the Sierra to the bottom of the continental shelf off shore amounts to more than 28,000 feet or about 5½ miles. The distance separating the two is about 230 miles. It is therefore not surprising that the area between is faulted in several places along lines generally parallel to these two boundaries.

A crucial chapter in this program related to the development of appropriate seismographs through which to locate the sources of landslips below ground and to determine the rate and conditions of propagation of the elastic waves caused thereby. Since we were operating within the zone of the earth movements themselves, the distance from epicenter to instrument and the corresponding time interval of shock transmission would be so short as to require instruments of high precision for their measurement, otherwise large errors were inevitable. This situation posed a new problem in earthquake observation requiring serious study and considerable experimental skill.

Mr. Wood associated with himself for this study the best experience available. A conference was called in Pasadena which was attended by Dr. Anderson of this Committee, by the late Mr. F. G. Pease of the Mount Wilson Observatory, by Dr. Frank Wenner of the United States Bureau of Standards, by Professor Arnold Romberg, then of the University of Hawaii, by Professor R. A. Millikan and several of his assistants from the California Institute of Technology. This group included the most expert information available in the several domains of physical experiment which must be brought together to produce a successful instrument of the scope contemplated in the plan. Several days were given to the consideration of the problem by this assembled group and different phases of it were assigned to individual members of the group for further study. The studies resulted toward the end of 1922 in a recording seismograph (afterward called the Wood-Anderson seismograph) based upon an entirely new principle. It is mentioned in the annual report of the Advisory Committee for the year 1923 in the following enthusiastic terms: "It is a great pleasure to report that this work has been crowned with most remarkable success. An instrument of the utmost simplicity, adaptable to the measurement of any one of the three components of movement and capable of detecting short waves as well as long ones, has been perfected in a remarkably short time." Two of these instruments were placed in service in Pasadena in January 1923 and from that time forward registration of local earth movements in southern California has been practically continuous.

Two other activities were inaugurated by the Committee at this time which were concerned primarily with hypothetical conditions upon which evidence was sought which, if decisive, would materially affect the direction and progress of our investigation. The first was concerned with the general problem of crustal creep. In the great San Francisco earthquake of 1906 there was a northwest and southeast horizontal slip along the San Andreas fault amounting in places to more than 20 feet and extending for perhaps 175 miles along the fault. Notwithstanding earlier surveys of the region it was by no means clear after the slip whether the eastern area had moved southward, or the western area northward, or both movements had occurred sharing the total slip between them. Observations of latitude undertaken at Ukiah, California, seemed to indicate that the Observatory and so the area to the east of the San Andreas fault was creeping northward. Certain observations at Lick Observatory, south of San Francisco, and also on the eastern side of the fault, appeared to confirm this. If Reid's theory is correct, that an earthquake results when crustal masses are gradually strained to the point of rupture, then a slow creep of a local crustal area becomes of the highest importance for it will almost certainly lead eventually to rupture and earthquake.

It was accordingly arranged that the latitude station at Ukiah, the operation of which was interrupted by the war, should be reoccupied by the Coast and Geodetic Survey and continuous observations for crustal creep be made there. Lick Observatory also agreed, in so far as such measurements were compatible with its regular astronomical program, to continue observations in the same sense.

Astronomical determinations of positions on the earth may be in error by as much as one meter. It was therefore only through continuation of such measurements over a long period of time that crustal creep could be established by these means.

The last of this series of correlated efforts of the Committee (in 1922) to delve into the problem of crustal adjustments in a region in which such adjustments were plainly indicated by frequent slips concerned itself with the problem of crustal loading, commonly called by geologists "isostatic adjustment." Very briefly this thought grew out of the gravity measurements carried out by the Coast and Geodetic Survey under the direction of J. M. Hayford in the early years of the century. Hayford's general conclusion was that the solid crust of the earth rested upon a deeper portion which was plastic enough to yield, very gradually, under load and that when such vertical adjustments had attained equilibrium the loading should be constant at all points above a certain isostatic level. If they were not complete or new shifts in the loading occurred then measurements of gravity at different points should reveal differences of mass or loading with accompanying vertical stresses tending always toward eventual adjustment. And so, just as we sought to establish by astronomical measurement local horizontal movement or creep, so also through measurements of gravity by the pendulum or otherwise it was thought to learn something of the distribution of local differences of mass below ground and the possibility of vertical stresses which conceivably might also cause displacement at the surface and earthquake.

The development of a more convenient instrument for gravity measurements was first intrusted to Dr. E. D. Williamson of the Carnegie Geophysical Laboratory, but its progress was interrupted through Dr. Williamson's untimely death. In point of fact, it is only during recent years that success has been attained[3] in providing a compact and portable instrument capable of measuring in a few moments the gravitational constant in any accessible place with an accuracy of approximately one part in a million.

Thus was inaugurated in the year 1922 a rather thoroughgoing study of earth movements and of the forces which participate therein, more extended perhaps than had ever been undertaken before by a single agency or coordinated group of agencies. It was not a problem to be finished and determined within any limited time, but rather a problem of gradual approach through continuing observation and deduction to a proper appraisal of great terrestrial forces and their relationship one to another and to human welfare.

---

[3] F. E. Wright and J. L. England.

Perhaps it is pertinent here to emphasize once again just wherein the present plan differed from others. I quote from the Annual Report of the Advisory Committee for the year 1923:[4]

> In reviewing the advances made during 1923 toward the study of California earth-movements, with which the Advisory Committee in Seismology has been charged, it is profitable to recall the fact that the procedure recommended by the committee and adopted by the President and Trustees of the Institution differs from the one usually followed in the pursuit of seismologic research. In the judgment of the committee, a single observing station, however carefully located, completely equipped, or competently served, would by itself yield little more than a routine record of local or distant disturbances. It is rather through a comprehensive study of surface features, of geologic structure, of the physical, particularly the elastic, properties of the rocks, of the distribution, direction, and magnitude of the tremors and displacements which occur and their relation to permanent zones of structural weakness, that an understanding of these movements and their proximate causes is to be sought. Of all these things the conventional seismograph, even of stations of the first class, can tell us only a part; of local earthquakes of short period, such as the California earth-movements for the most part have been shown to be, it can tell us extremely little.
>
> It is for these reasons that your committee has departed from the practice which has been developed in other countries where earth-movements are studied, and sought the cooperation of those agencies which, more competently than any others, are able to furnish precise information about the surface and the structure of the portion of the earth's crust which it is proposed to study.

As has been stated above, the first seismographs of the short-period, high-magnification, Wood-Anderson type, for the measurement of local earthquakes, were set up at the end of 1922. Early in 1924 these were replaced by newer instruments of the same type but of an improved construction through which competent data began to be obtained and the sources of these local shocks to be located and mapped. The early epicenters thus determined did not fall uniformly upon the geologically determined fault lines already established. Possibly half of them were scattered about other areas, indicating other regions of crustal weakness—perhaps faults hitherto concealed by surface conditions common in the desert areas or perhaps of quite other origin. This was probably the most suggestive observation of these early years.

---

[4] Carnegie Inst. of Wash. Year Book No. 22, p. 362.

The fault-zone map of 1923 had depended entirely upon the geological appraisal of surface indications. Now we were beginning to get supplementary subsurface evidence which could not fail to add in a most important degree to our knowledge of these phenomena.

In 1924 also the first chain of monuments and primary triangulation was completed by the United States Coast and Geodetic Survey. It extended from Mounts Lola and Round Top in Nevada westward across the northern fault zones to San Francisco Bay, continuing thence southward along the coast to the Mexican boundary and thence again eastward into the Colorado River valley. To this system a spur was added northward from San Francisco Bay to include the latitude station at Ukiah, which was engaged upon the continuous observation of local latitude through which crustal creep, if present in that region, was to be established. This system of triangles crossed the known fault zones northeast of San Francisco in several places and again the more intricate series of faults known to exist but so far not well located in southern California. Through this chain of triangles it was hoped to evaluate in direction and magnitude any appreciable earth movements which might occur in the zones traversed by the triangulation.

This primary survey proved its usefulness in the following year when the destructive Santa Barbara earthquake occurred within the region under observation. It was possible to prove after the earthquake that the disturbance did not proceed from the major San Andreas fault some 30 miles away to the northeast, but was entirely a local disturbance associated with small faults about the Santa Ynez Mountains. Although the movement was violent, damage was almost confined to the main street of Santa Barbara city and even here measurable permanent displacements of the ground were not observed. The results of the year were instructive in that they served to show that the earth movements of the west coast region are not necessarily confined to the great known faults, although the threat is perhaps greatest there. They also served to add emphasis to the importance of the observations of the seismographs which had already located points of stress release more or less remote from these fault zones.

In this year (1925) preparation was also made for the study of possible vertical movements resulting from these crustal stresses. Geological observations had plainly indicated the importance of the vertical movements of the past, seen for example in the great Owens Valley fault forming the eastern boundary of the Sierra Nevada Mountains, where a vertical displacement of some thousands of feet is indicated. A beginning of precise leveling was made in this year by the United States Coast and Geodetic Survey along the line of the Santa Fe Railway from Los Angeles over the Cajon Pass to the state boundary. The choice of this particular route was partly a matter of convenience and partly because a line of roadbed levels from earlier years was available for preliminary comparison for the purpose of detecting any vertical movements which might have occurred in the preceding fourteen-year interval. The line of the railway crosses the San Andreas fault near Cajon Pass.

There were also in existence two earlier determinations of the height of Mount Whitney which made it desirable to redetermine the elevation of this landmark in order to ascertain whether the mountain mass of the Sierra was changing with respect to neighboring areas separated from it by known fault zones. Accordingly precise levels were measured across the eastern fault zone (Owens Valley) and up the mountain slope to an elevation of 8000 feet, where, because of the lateness of the season and the destruction of the mountain trail, this effort was interrupted.

In this year also advantage was taken of an appropriation of $5000 to the Advisory Committee by the Carnegie Corporation of New York for seismological publication, to print in the *Bulletin of the Seismological Society of America* (March 1925) a "Description and Theory of the Torsion Seismometer," the new instrument for the study of local earthquakes to which reference has already been made.

In the years 1925 and 1926 tangible progress was indicated in several of the directions in which the Committee's efforts had been exerted. First of all there appeared in this year (1926) the first detailed geological report upon "The San Andreas Rift and Some Other Active Faults in the Desert Region of Southeast California" by Dr. L. F. Noble of the United States Geological Survey. Dr. Noble traced this major fault along the southern border of Mohave Desert near the base of the San Gabriel range, southeastward to

Cajon Pass, thence along the southern base of the San Bernardino Range through San Gorgonio Pass into the Salton Sink. Along the rift a strip averaging 6 miles in width was mapped and studied in detail.

This report, somewhat preliminary in character, was printed in the Report of the Advisory Committee[5] pending full publication as a Bulletin of the United States Geological Survey later. Dr. Noble has this to say in characterization of the San Andreas Rift in this active region of southern California:

> The belt of branching, interlacing, and roughly parallel faults that borders the main San Andreas fault and constitutes the San Andreas fault-zone exhibits a structure so complex that a simple statement concerning it is difficult to make. In general, the zone is a fault mosaic of elongated sliver-like blocks whose longer axes trend parallel with the strike of the main fault, but at many places the rock masses are so intricately shattered and different formations so mixed together that it is impossible to map them or to determine their relations and age. The evidence of the nature of the fault-movements is somewhat conflicting; the dominant structure is a sort of slicing which appears to be mainly the result of horizontal shear along the San Andreas fault. . . . Normal faults occur at a number of places in the fault-zone. Some are of considerable magnitude and have high scarps. Graben structures are common in some parts of the zone. The normal faulting, however, is a relatively unimportant feature, probably the result of vertical adjustments within this zone of profound weakness.
>
> One noteworthy feature of the fault-zone is the thorough crushing and shattering of the harder rocks everywhere along the main fault, usually through a belt several hundred yards wide. In this belt the rocks are at most places so ground up that they are practically gouge. At many places the rocks are thoroughly decomposed. Granite, for example, at most places along the fault is so rotten that when wet it becomes plastic like clay.

About the structurally important Garlock fault he adds the following (page 424):

> The Garlock fault apparently marks a significant structural and topographic boundary in the region. At the fault, or not far north of it, all the north-south structures which characterize that part of the Great Basin lying between the Sierra Nevada and Death Valley come to an abrupt end. . . . South of the fault the geologic structures do not exhibit the parallel orientation that they do north of the fault, nor do they trend consistently in any one direction.

In the same year (1925–1926) we were still further encouraged in our efforts by an arrangement for closer cooperation with California Institute of Technology at Pasadena. It had been our hope from the beginning that a research program so closely related to the welfare of the west coast region might attract tangible support from the community served. Now that the program began to be better understood and the sincerity of our purpose to be correctly appraised, and perhaps also after the reminder furnished by the destructive earthquake at Santa Barbara (1925), a real and much more general interest began to be shown in the effort to learn all that we could learn about the west coast earth movements and their causes. As on a number of other occasions where a convincing scientific need had been shown, the late Dr. George E. Hale took the initiative in securing from California Institute of Technology a promise to build a suitable laboratory in Pasadena to house the seismological research staff and interested students, and to provide piers, tunnels, and other special equipment upon granite foundations for mounting central-station instruments and for general research purposes in seismology.

---

[5] Carnegie Inst. Wash. Year Book No. 25, p. 415.

In pursuit of this plan a tract of three acres located on a granite hillside just west of the City of Pasadena has been purchased and ground broken for a laboratory building which is already in process of construction and is expected to be occupied soon after January 1, 1926.

The ground floor will provide a seismograph room with six piers built directly upon the underlying granite, a small laboratory room with two additional piers for experimental purposes, a photographic room, a radio room, and a large storage and file room. At the back of the building two tunnels will be driven into the granite rock composing the hillside, one to a depth of 36 feet, the other 30 feet. The two tunnels will enter from opposite ends of the building and will be connected by a covered passage the entire length of the building (78 feet). This passageway with tunnels at right angles to it at each end will afford excellent opportunity for tilt experiments in two dimensions which have long formed a part of the Committee's plan.

On the second floor are two laboratories, four offices, a large drafting and measuring room, and an attractive lobby at the main entrance, which is on this floor.

The entire building is of concrete, about 78 by 40 feet in floor area exclusive of tunnels and passageway, has a tile roof, and is of nearly earthquake-proof construction.[6]

Arrangements were also made in this year for a number of associated outlying stations mounting instruments similar to those in the central (Pasadena) laboratory through which, by triangulation and otherwise, the sources of the earth movements occurring in the region might be accurately located. The first of these stations was provided by the city of Riverside, about 50 miles east of Pasadena, the second by the Scripps Oceanographic Laboratory at La Jolla, to the south of the main station, the third in the Mount Wilson Observatory of the Carnegie Institution at the top of Mount Wilson. Following the Santa Barbara earthquake a number of public-spirited citizens provided an additional station upon the grounds of the Natural History Museum in that city. Of these stations Riverside, Pasadena, and Mount Wilson are provided with piers in direct contact with the underlying granite. The other two (La Jolla and Santa Barbara) rest upon water-laid sediments more or less consolidated. Altogether they have formed a fairly satisfactory group for beginning detailed studies of the origin and progress of earth movements in the region, though additional stations to the northeast and southwest are contemplated in the plan. The new Wood-Anderson seismographs were mounted in these stations to register the two horizontal components of movement and a third pier was provided to receive a vertical-component instrument as soon as it should be perfected.

---

[6] Carnegie Inst. Wash. Year Book No. 24, p. 378.

With actual stations provided for and with the advent of sensitive and precise seismometers a major technical difficulty, recognized from the beginning, began to receive more attention. The determination of the point of origin of an earth tremor presupposes some accurate means of determining the time of transit of the elastic wave from source to instrument, and the rate of propagation of such impulses in the region under investigation. This rate of transmission can sometimes be determined from artificial explosions of known point of origin and distance from the recording instrument.

From the beginning of experimental registration in 1923 to 1930 blasts were recorded from nine different sources at nineteen different distances between blast and seismometer station ranging from about 4 km up to 195 km. The blasts have differed greatly in magnitude and the records differ considerably in quality and the degree of control of time, etc. Notwithstanding this, an intensive study of the whole assemblage carried out recently has brought out very consistent results with useful determinations of phase and wave speed. This group of records and the results have been published separately.[7]

The second factor entering into this determination is of course the accuracy of the travel-time record upon the seismogram itself. Such records are written photographically upon a rotating drum advancing at a constant rate, in this case 60 mm per minute. Precision depends upon the accurate and uniform rotation of the drum and the uniformity of the time signals impressed upon it. To provide a continuous record throughout an earthquake of more or less severity such a mechanism must possess sturdy qualities which are sometimes difficult to obtain. Time signals from a pendulum clock, for example, which might stop at the critical moment, would therefore be of limited usefulness.

This problem, as may be imagined, has given some difficulty and has required several years before a completely successful solution was reached. At present determinations of time are based upon automatically recorded radio time signals sent out from Washington several times daily by the Navy Department. Intervals between these time signals depend for their precision upon the mechanical arrangements within the seismograph set-up itself. At present an electrically driven tuning fork provides the subdivisions of time between signals, and synchronous motors of special slow-moving type provide the uniform rotation of the drum. In recent years, since a tenth second has become important in the time determinations, a parallel record of these minute marks alongside the dots and dashes of a powerful commercial broadcasting station is carried on another drum at all the associated stations. In this way identical signals (dots and dashes) check the time marks upon all the seismograms from the widely separated stations in service.

Two other events of the year 1925 are noteworthy as indications of the progress in the study of American seismology which the Advisory Committee had set for its major purpose. The first of these was the founding of the Jesuit Seismological Association under the presidency of Professor J. B. Macelwane, S. J., of St. Louis. A number of Jesuit stations had been in existence in the Jesuit colleges of the United States and in other countries, but they had not, up to this time, concerned themselves with seismological research so much as with the public interest in earthquakes and with the technical education of students. The new organization on the contrary stated its purposes explicitly:

---

[7] Harry O. Wood and Charles F. Richter, *Bull. Seismol. Soc. Amer.*, vol. 21, March 1931.

> (1) To provide for uniform standards of measurement and interpretation of seismograph records.
>
> (2) To encourage seismologic research.

Professor Macelwane himself is one of the foremost students of seismology in this country.

The second event indicating the public interest in our efforts on the west coast was the establishment of a Committee on Building for Safety Against Earthquakes organized by the Seismological Society of America in cooperation with the Society of American Civil Engineers, the American Institute of Architects, the National Board of Fire Underwriters, and several other associated organizations. This Committee took its task most seriously and during the years 1925 and 1926 published a number of articles on engineering problems in relation to earth movements in California. The work of this Committee has also been extended to the consideration of practical methods of enforcing better construction. Up to this time building codes had taken no account of the stresses which may be set up by an earthquake. It became necessary therefore not only to establish the magnitude of the stresses and the general features of building design necessary to resist them, but to write into the building codes these detailed specifications together with provision for their enforcement.

The year 1927 and the early part of 1928 provided something of a breathing space in the activities of the Advisory Committee. The energies of the laboratory group were largely occupied with the routine operations of equipping the new central laboratory in Pasadena and the four outlying stations associated with it. The surface triangulation work of the United States Coast and Geodetic Survey likewise paused in its progress for a period of nearly two years during which its energies were largely directed toward a general readjustment of all the closing errors in the primary triangulation of the western half of the United States.[8] This was in itself a task of great magnitude.

---

[8] See A Comparison of Old and New Triangulation in California, by Dr. William Bowie, chief of Division of Geodesy of the United States Coast and Geodetic Survey; obtainable from the Superintendent of Documents, Government Printing Office, Washington, D. C.

Advance sheets from this readjustment of the triangulation net were of course of the greatest interest to us. From them it appeared that the major system of triangles, which had been established during the preceding five years, as a reference system for the evaluation of earth movements in California as they may occur, was securely anchored at its two ends (1) near Reno, Nevada, (2) near American Peak in Colorado River valley. These regions had suffered no displacement since the earlier surveys some thirty years before. Likewise in southern California displacements were for the most part negligibly small, but in the vicinity of Point Reyes and Ross Mountain toward the northern end of the San Andreas fault considerable displacements were indicated, as might be expected from the fact that this was the first general readjustment since the San Francisco earthquake of 1906. Ground movements of several feet in this region were discovered and accurately indicated in direction and magnitude.

Of the operations in progress during this year, the determination of the height of Mount Whitney (begun in 1925) was completed without discovering any change of altitude greater than the errors of measurement in precise leveling.

During this interval also additional lines of precise leveling for reference in the case of vertical earth movements were run in several of the faulted areas in southern California. Indeed, one of these leading from El Centro served its purpose in the following year when a sharp earthquake in that region produced vertical displacements of several inches, the course of which could be accurately followed from this line of levels.

One further advance planned by the Advisory Committee for the purpose of obtaining a more intimate knowledge of the actual orientation of earth movements within and immediately adjoining a fault zone was inaugurated in this year (1928) and continued at intervals during several subsequent years.

The network of primary triangulation, which has been extended over those regions of California where earth movements are known to occur, is perforce a scale of large units. The monuments are from 10 to 50 miles apart and consequently very few of them happen to fall sufficiently close to known faults to reveal local displacements near the fault trace or the distribution of the strains produced in the adjacent ground before or during fault movement.

Accordingly the Committee prepared a plan to supplement the primary triangulation with lines of closely spaced monuments perpendicular to certain of the more important faults and extending for 4 or 5 miles on either side of these. Several such lines were contemplated at appropriate places both in northern and southern California in order to determine the kind and amount of movement now going on within its chief fault zones. With a suitable number of these closely spaced monuments, aligned perpendicular to the fault trace and accurately determined in position both horizontally and vertically, both the character and amount of displacement immediately at the fault and at various distances away from it could be determined, whether the movement were in the nature of a slow creep over a period of years or a sudden release of strain. It is hoped that this plan will provide even closer control of surface movements in faulted regions than has been attained before or elsewhere.

It is perhaps opportune to point out here that the severe earthquake in Santa Barbara in 1925 provided an atmosphere more favorable to further study of California earth movements than had prevailed before. The San Francisco Bay region in particular, which had learned the taste of catastrophe in 1906, became suddenly sensitive to the earthquake hazard and under the advice of Dr. Bailey Willis, of this Committee, made provision for four seismograph stations appropriately placed about San Francisco Bay through which the local earth movements might be kept under continuous observation there. These four stations were equipped and operated like the stations immediately associated with the Pasadena laboratory and provision was made for the exchange of records and suggestive information between the two groups. Although there is a considerable gap between the station at Santa Barbara in the southern group and the nearest of the northern California stations at Mount Hamilton (something over 300 miles), nevertheless the collaboration, continuous since 1928, has provided the basis for a much broader analysis of the entire group of west coast movements than could have been obtained from either group separately. These northern stations have been in charge of Professor Perry Byerly of the University of California and are located at Berkeley, Mount Hamilton, Palo Alto, and Golden Gate Park, San Francisco.

In the following year (1929) the ground covered by the seismograph stations of the southern group was extended northeastward by the addition of stations at Haiwee and Tinemaha along the line of the Los Angeles aqueduct in Owens Valley. Shelters and piers for these stations, and attendance, were provided by the Los Angles Water Company, the instrumental equipment and its maintenance by Carnegie Institution. These were the first stations in the California group which were located to the east of the Sierra Nevada range and therefore beyond the great group of faults associated with it.

With the completion of the central laboratory at Pasadena and the equipment of the group of outlying stations south and north, the plan which had been developed by the Advisory Committee was in full operation, limited only by the number of stations available. It therefore seemed appropriate to pause in the year 1929 for a reappraisal of the various approaches to the west coast earthquake problem which the Advisory Committee in Seismology had thus far made at the bidding of the President and Trustees of the Carnegie Institution.

It should perhaps be explained that the Advisory Committee in Seismology is country-wide in the distribution of its membership. Accordingly its operations have been carried on during the eight years of its existence by subgroups meeting occasionally, by correspondence, and by the more or less consecutive efforts of the chairman. For purposes of the proposed review a conference of the full Committee was called to meet at the Seismological Laboratory in Pasadena in October 1929. To this conference were bidden, beside the members of the Committee, two of the most distinguished students of seismology abroad, Messrs. Harold Jeffreys of Cambridge, England, and Beno Gutenberg of Frankfurt a.M., Germany, also the president of the Jesuit Seismological Association, Dr. J. B. Macelwane, S.J., and certain representatives of other institutions known to be interested. There was laid before this group the full record of the efforts of this committee for the advancement of American seismology. These included the fault map of California, the contour map of the ocean floor along the adjacent coast, and the triangulation net and precise levels undertaken by the United States Coast and Geodetic Survey throughout the west coast region in which earth movements were known to occur. It was also pointed out that this system is now being supplemented by local triangulation or lines of monuments, averaging one monument to the mile, crossing the great fault zones at right angles in selected regions north of San Francisco and south of Los Angeles, for the purpose of determining the distribution of strain or crustal creep in the immediate neighborhood of the fault itself. There is certainly no region in the United States in which provision for the detection and measurement of horizontal and vertical earth movements through the location of bench marks at critical points is more completely developed than in the earthquake zone of California. By no means least of these activities is the splendid impetus given to the geological investigation and mapping of fault-zone regions by the work of Dr. Levi F. Noble of the United States Geological Survey upon the southern section of the San Andreas fault, supplemented locally in other sections by the geological faculties and students at the University of California and Stanford University.

There were also laid before this assembly at Pasadena the complete records of earth movements, large and small, as written on the seismograms taken from the various instruments in operation since the first installation in Mount Wilson Observatory in 1923. All the new instruments developed during this period were exhibited in operation and opportunity was given to visit one of the outlying stations (Santa Barbara). The members of the Committee and its guests were also given opportunity to travel along the surface outcrop of the San Andreas fault for a distance of some 60 miles northwestward from San Bernardino, in order to learn at first hand the appearance of these great zones of movement and the various provisions for their continuous observation. The program of informal conferences on matters of common interest lasted for two weeks, the opportunity for the interchange of ideas and suggestions being quite unlimited.

In consequence of this rather intimate and detailed scrutiny of the results of its efforts and the enthusiastic recommendations of the distinguished guests of the Committee, the Advisory Committee felt some confidence in the soundness of its progress, and in the business meeting of the Committee, with which the gathering at Pasadena closed, it was voted to recommend the continuance of the present program without essential modification for a period of years.

Reporting in the following year upon this conference in Pasadena in the autumn of 1929, the chairman of the Advisory Committee in Seismology has this to say upon the situation as developed there:

> The meeting of the Advisory Committee in Seismology, held in Pasadena in the summer of 1929 for the specific purpose of reviewing the progress of its work during the eight years of its administrative responsibility, brought clearly into the foreground a number of rather satisfactory conclusions:
>
> (1) The plan which had been recommended to the Carnegie Institution of Washington by the Advisory Committee in 1921, and which has been developed as a "tentative" project under its direction since that time, has received the full endorsement of the group of specialists which had been gathered there to review in detail the work so far accomplished.
>
> (2) Progress during the eight years in the various directions set forth in the plan was also found to be satisfactory. Not only have new and extremely sensitive instruments been developed [where none existed before] for recording local shocks, but some thousands of earth movements have already been satisfactorily recorded. Indeed, it was said, informally, by one of the visitors, that a greater body of competent material for the study of such local movements has been gathered here than can be found elsewhere, except possibly in Japan.
>
> (3) That upon the initiative of the Committee, extensive geologic and topographic studies of the region have been made by cooperating agencies which cannot fail to add greatly to our knowledge of these movements within the coastal earthquake zone. These agencies are national in scope and deserve particular mention by name: The Geological Departments of University of California, of California Institute of Technology, and of Stanford University; the U. S. Coast and Geodetic Survey, the U. S. Geological Survey and the U. S. Navy Department; and the Seismological Society of America.
>
> (4) It was also a matter of gratification to the Committee that those cooperating agencies have become so much interested in the program of studies which has been undertaken that pertinent additional work has been contributed by many of them in the conviction that the study of the earth movements of this region is of the highest importance, both from the point of view of geologic evolution and of human welfare.
>
> I think it may fairly be said that the close of the Conference held in Pasadena left a firm conviction in the minds of all who were present, both members of the Committee and distinguished guests, that the undertaking which has hitherto been described as "tentative" has come to deserve a more permanent status.
>
> Whether this be the logic of the situation or not, cooperation has been and remains of the very essence of it, and it is upon the collaboration between these agencies, each of which contributes a vital part, but no one of which comprehends all of the activities here contemplated, that its future must depend. It is in this sense a particular pleasure to announce that the California Institute of Technology last year (1930) established a chair of Geophysics and called to it the foremost of the continental seismologists, Professor Beno Gutenberg, who immediately joined his own activities to ours and made his headquarters at the Pasadena Seismological Laboratory.

Progress from this point on was attended with considerable enthusiasm following naturally upon the assurance of greater permanence in the relationship of the cooperating groups, the competent and altogether satisfactory performance of the various instruments which had been developed, and the assurance of continuity in the observation of earth movements in the region being studied. Also with the accumulation of a great body of earthquake records for study (something like 1000 earth tremors of varying intensity were being clearly recorded annually), and the opportunity now afforded to analyze and interpret these data in appropriate perspective, it was determined to make a beginning of publication of the more important data registered. Accordingly, beginning with the calendar year 1931 records of all important earthquakes, both distant and local, were circulated monthly to all observatories requesting them. Thus the Seismological Laboratory at Pasadena took its place with the great seismological laboratories of the world following upon the assurance of the distinguished guests attending the conference in 1929 that already the Laboratory was in possession of a greater volume of competent data upon local earthquakes for study than had been accumulated elsewhere.

Notwithstanding this confidence in the data already accumulated and in the instruments which were daily adding to them, it was not possible until this year (1931) to announce the development of a thoroughly satisfactory vertical-component seismometer of short period. Such an instrument was successfully developed by Dr. Hugo Benioff of the Pasadena Laboratory staff and responded satisfactorily to all the tests applied to it. An instrument of this type was added in this year to complete the equipment of all the stations of the southern California group, now seven in number. It happens also that this instrument proved equally effective for long-period registration (distant earthquake) and instruments of this type were furnished in this year to the Department of Terrestrial Magnetism of the Carnegie Institution of Washington for such use at Huancayo, Peru.

It is appropriate also to record the fact that in this year (1931) the technique of time registration for earthquake records, which has already been mentioned as a source of considerable trouble and anxiety in earlier years, appears to have been developed to a satisfactory point.

It is worth noting in connection with the vertical-component seismograph, which was the first to be provided with galvanometric registration, that it proved entirely possible to operate from a single vertical-component inertia-mass two different vertical-component seismometers adapted to the separate registration of both far and near shocks. Also the increased magnification, which in the case of the short-period assembly amounted to about 5000, has led to the recognition of many disturbances which otherwise would have been passed by as doubtful. At the same time there has not been so large an increase in the number of recorded shocks as might be expected from the use of this higher magnification. Apparently increase of magnification does not necessarily bring to light more and more true earthquakes.

The successful operation of the Benioff vertical-component seismometer (1931) led to the design and construction of a horizontal-component instrument having similar properties. The constants of this horizontal seismograph assembly were arranged to be identical with those of the vertical assembly so that it became possible for the first time to make direct comparisons between seismographic records of vertical and horizontal earth movements. These instruments represented the first departure from the torsion-type instrument of 1923, over which they possessed advantages both in sensitiveness and in adaptability. They probably represent the greatest advance in earthquake recording which has been attained up to this time. Recognition of this fact has doubtless contributed to their extensive adoption in later years elsewhere.

The year 1932 brought the completion of Dr. Noble's detailed study of the southernmost 80 miles of the San Andreas fault in California, to which reference has already been made. In it he directs attention to the view entertained previously by a number of geologists that the entire San Bernardino range north of the fault has been shifted many miles southeastward with reference to the San Gabriel range south of the fault. Of this he says: "No definite stratigraphic evidence of horizontal movement on such a scale was obtained. The evidences of horizontal movement indeed are contradictory." Nevertheless in his conclusion he states confidently that a vast amount of movement has taken place along the line of the present San Andreas fault throughout Tertiary and recent geologic time. It is appropriate to quote briefly from his concluding paragraph:

> The profound difference in the rocks on opposite sides of the San Andreas fault shows that the fault movements have been of great magnitude, but in the eighty-mile sector studied by the writer it is not possible exactly to date them by stratigraphic evidence. It seems unlikely that the line of the curiously straight and continuous master fault could have originated as anything but a shear zone on a nearly vertical plane. It is seemingly too straight over great distances to have originated as a normal fault or as a low-angled overthrust. As contrasted with the older Quaternary movements the recent movements seem very small. . . . . The writer's impression of the fault is one of gradually dying activity in late Quaternary time, as contrasted with violent activity in the early Quaternary to late Tertiary—yet this impression may be but testimony of the relative shortness of late Quaternary time.

It is hardly appropriate in this place to describe the many different types of apparatus which have been developed from time to time in the Seismological Laboratory for the more effective study of earth movements. These cover the entire range from apparatus of the highest sensitiveness for the detection of the smallest tremors to "strong-motion" instruments calculated to take over and record movements of catastrophic magnitude. Nearly all of these have found their place in the study of earth movements of various types in the California region. The idea of strong-motion instruments has been developed rather extensively in recent years by the Division of Terrestrial Magnetism and Seismology of the United States Coast and Geodetic Survey under the personal direction and enthusiastic interest of Captain N. H. Heck. A considerable number of these instruments, developed and built in Captain Heck's division, and operated by its personnel, are now in operation along the entire west coast area, in Montana, and elsewhere in regions assumed to be exposed to earthquake.

One development of the Seismological Laboratory is perhaps sufficiently novel and useful to receive mention here. It has been described as a "portable seismometer station," an assembly of apparatus in transportable form for the registration, at different selected points, of earthquakes, blasts, and similar disturbances. It is really a complete station equipment for receiving and recording all three components of earth movement, even including a radio-code receiving apparatus for checking with central laboratory records the time signals recorded upon the seismograms. Provision is also made for changing the photographic record sheets and performing all operations necessary for a field station. In use the vertical-component seismometer (Benioff type) is placed on the best ground or rock available and protected from wind and weather by a cylindrical metal housing with cover. This equipment mounted upon an automobile trailer is available for instant use in emergency and proved particularly valuable in recording the after-shocks of the great Long Beach earthquake of March 10, 1933, supplementing the records from the permanent and more distant stations.

The Long Beach earthquake, of which an extended study has been made and published elsewhere, afforded opportunity for observations directly related to the security of inhabitants and buildings. These observations were planned to determine the effect of after-shocks in high buildings from the basement upward and were conducted in cooperation with the Department of Civil Engineering of the California Institute of Technology. Four torsion seismometers from the Pasadena Laboratory were remodeled according to a design by Dr. Benioff and used by Professor Martel in several of the high buildings of Los Angeles, where they yielded engineering information of considerable value to architects and city engineers. It appeared from the records obtained that the movements at the top of some buildings were as much as three times as great as the corresponding ground movements and resonance effects were very pronounced in all the buildings tested.

The opportunity for detailed study of a major earthquake (Long Beach) and its after-shocks in a region well equipped with instruments of the highest precision and well monumented with accurate landmarks, both in vertical and horizontal dimensions, was the first of its kind to come within our program. It was due of course to the forethought of earlier years on the part of the Advisory Committee, of the Division of Geodesy of the United States Coast and Geodetic Survey, and of other cooperating agencies.

In point of magnitude it was not the largest earthquake which had occurred during the period of our activity in the west coast region. At least four others have been recorded which represented greater ground movement, but it happened to occur in a well-populated region in which, both from scientific interest and from the point of view of human welfare, our efforts had been concentrated. Loss of life was not large considering the concentration of population in Los Angeles and its neighborhood, but destruction of buildings was widespread, and the thought inevitably comes uppermost in the minds of witnesses of this destruction that if the earthquake had occurred four or five hours earlier in the day, the loss of life, particularly among the school children, must have been very great. The destruction of school buildings in Long Beach and adjacent cities in the Los Angeles plain was general and often complete.

This had the momentary effect of arousing and bringing out to the light of day the latent earthquake consciousness of the entire community, and efforts to forestall or limit similar destruction for the future became general. Even the City Council of Los Angeles, which had opposed all publicity and earthquake investigation in time past, now published a general warning that southern California was a region of moderate seismic activity, that occasional shocks might be expected in future as in the past, and that engineers, architects, builders, and others interested should make appropriate efforts to add building strength where needed both to new structures and to structures under repair with a view to preventing similar disasters in future. Even the Public Works Administration in Washington was moved to allot the sum of $40,000 for an intensive study of the effect of earthquakes upon high buildings, and the United States Coast and Geodetic Survey in collaboration with the Carnegie Institution of Washington, California Institute of Technology, and the Joint Technical Committee on Earthquake Protection of Southern California was authorized to undertake the task. Under this allotment new instruments were constructed in the Seismological Laboratory and upwards of 250 high buildings, elevated water tanks, bridges, dams, and other structures particularly sensitive to earth movements were studied in detail for the engineering information they could give.

Following upon the Long Beach earthquake the Advisory Committee asked the United States Coast and Geodetic Survey to repeat the triangulation from Newport Beach to Lucerne Valley in order to pick up and determine the location of any permanent displacements which may have occurred, but none was found. Subsequently this line of triangles was continued on from Lucerne Valley to Needles in order to strengthen the network now spread widely over the southern California earthquake zone in order to establish such displacements when they occur.

In this year 1935 also a very intensive leveling survey was made at the request of the chairman of the Committee, with monuments even more closely placed than heretofore in order to determine with microscopic care both movements and trends along the main fault lines of southern California. These lines cross the San Andreas fault at Maricopa, Bailey's Ranch, Palmdale, Cajon, and White Water. Another line crosses the double San Jacinto fault in the vicinity of Armada. Still another crosses the Whittier fault in the vicinity of Brea, and finally the Inglewood fault, which was the seat of the Long Beach earthquake, was similarly documented at Inglewood. Each of those lines extends for about 5 miles out from the fault in either direction with monuments spaced at intervals of 100 feet for the first mile, 200 feet for the second mile, 300 feet for the third mile, 400 feet for the fourth mile, and 500 feet for the fifth mile. It is interesting to note that the Inglewood fault was still active in 1935 and small shocks were frequently reported from that source in that year.

In 1934 and 1935 a new type of earthquake began to be recognized upon the records of the teleseismic instruments. These were the so-called deep-focus earthquakes which appeared to originate in the south Pacific Ocean both east and west and in the Bering Sea region. No earthquakes of this type have been identified upon the west coast of the United States.

These earthquakes, characterized by a dominant vertical component, are peculiar in that they seem to originate at a depth within the earth's crust of from 300 to 700 km. At this depth the earth's texture is generally thought to be plastic and continuous and no adequate explanation has yet been found of any physical discontinuity which could give rise to earth shocks of this magnitude.

Such earthquakes began to register upon the seismographs of the Seismological Laboratory as early as 1931, and in 1934, 179, or about one-third of the distant shocks, were thought to be of this origin. During this period these deep-focus earthquakes have been generally observed elsewhere also and numerous publications have sought to provide some physical explanation for their occurrence but thus far without conspicuous success.

With the year 1936 the particular chapter in the development of American seismology in which the Advisory Committee had been charged with an administrative rôle came to an end. The major item and central fixture in its plans for the observation and study of earth movements had been the establishment, in the southern earthquake zone on the west coast, of the Seismological Laboratory and associated stations for the measurement and study of local earthquakes in that region. Among the institutions cooperating with the Advisory Committee in this development California Institute of Technology had shown a continually increasing interest in the problem and, because of its location within the earthquake zone, had come to share the responsibility for its progress.

In 1926 the building now occupied by the Seismological Laboratory was erected by California Institute and turned over to the Advisory Committee for use in these studies. In 1931 Professor Beno Gutenberg, a leading German seismologist, was appointed by California Institute to its chair of geophysics, and became a member of the staff of the Seismological Laboratory. Since that time also the financial burden, hitherto borne by Carnegie Institution of Washington, has been shared by California Institute. It was therefore quite natural that in 1936 the management of the enterprise should gradually come to center in California Institute, a local institution as opposed to Carnegie Institution with its headquarters in Washington, D. C. The agreement was that joint support should be continued for a period of five years with gradually diminishing participation by Carnegie Institution, and that the management should pass into the hands of California Institute of Technology from 1936.

Thus the "adventure in collaboration," inaugurated by President Merriam in 1921, after a successful record of fifteen years, during which the interest of the collaborating groups appeared to be a continually increasing one, with the enterprise in full fruition, was turned over to other hands in which its future seems well assured, not only for the advancement of research in seismology, but for the very considerable benefit and enlightenment of an increasing population in the southwest country, which has always to reckon with the possibility of earthquake disturbance.

The best record of the results thus far accomplished by the Advisory Committee in Seismology and associated agencies is contained in its publications, a list of which is appended for the benefit of those who may be interested to follow in further detail the more recent intensive studies of earth movements in the United States.

PUBLICATIONS

Hugo Benioff, Operating Frequency of Regenerative Oscillatory Systems, *Proc. Inst. Radio Engineers*, vol. 19, no. 7, pp. 1274–1277, July 1931.

Harry O. Wood and Charles F. Richter, Recent Earthquakes near Whittier, California, *Bull. Seismol. Soc. Amer.*, vol. 21, no. 3, pp. 183–203, September 1931.

B. Gutenberg and C. F. Richter, On Supposed Discontinuities in the Mantle of the Earth, *Bull. Seismol. Soc. Amer.*, vol. 21, no. 3, pp. 216–222, September 1931.

B. Gutenberg and C. F. Richter, Pseudoseisms caused by Abnormal Audibility of Gunfire in California, *Gerlands Beitr. z. Geophysik*, vol. 31, pp. 155–157, 1931.

Harry O. Wood and Frank Neumann, Modified Mercalli Intensity Scale of 1931, *Bull. Seismol. Soc. Amer.*, vol. 21, no. 4, pp. 278–283, December 1931.

C. F. Richter, Earthquake of January 28, 1931, *Bull. Seismol. Soc. Amer.*, vol. 21, no. 4, p. 284, December 1931.

B. Gutenberg, Travel Time Curves at Small Distances, and Wave Velocities in Southern California, *Gerlands Beitr. z. Geophysik*, vol. 35, pp. 6–45, 1932.

B. Gutenberg, C. F. Richter, and H. O. Wood, The Earthquake in Santa Monica Bay, California, on August 30, 1930, *Bull. Seismol. Soc. Amer.*, vol. 22, no. 2, pp. 138–154, June 1932.

Hugo Benioff, A New Vertical Seismograph, *Bull. Seismol. Soc. Amer.*, vol. 22, no. 2, pp. 155–169, June 1932.

At a meeting of the Seismological Society of America held on April 8, 1933, at the University of California at Los Angeles, informal presentations of the following papers were made by members of the Seismological Laboratory staff: "Recent Developments in Electromagnetic Seismographs," by Hugo Benioff; "Advantages of Using Geocentric Latitude in Calculating Distances," by B. Gutenberg and C. F. Richter; "Preliminary Results on Three Recent Great Earthquakes," by B. Gutenberg and C. F. Richter; "An Instrumental Scale for Earthquake Magnitude," by C. F. Richter.

Papers were submitted for presentation at the Fifth Pacific Science Congress at Vancouver, in June 1933, as follows: "Seismological Research in Southern California," by Harry O. Wood; "A New Electromagnetic Seismograph," by Hugo Benioff.

B. Gutenberg, Harry O. Wood, and John P. Buwalda, Experiments Testing Seismographic Methods for Determining Crustal Structure, *Bull. Seismol. Soc. Amer.*, vol. 22, no. 3, September 1932.

Harry O. Wood, Preliminary Report on the Long Beach Earthquake, *Bull. Seismol. Soc. Amer.*, vol. 23, no. 2, April 1933.

Harry O. Wood and Charles F. Richter, A Second Study of Blasting Recorded in Southern California, with a note by Hugo Benioff, *Bull. Seismol. Soc. Amer.*, vol. 23, p. 3, 1933.

B. Gutenberg, Über Erdbeben mit Herdtiefen von Mehreren Hundert Kilometer, *Geol. Rundschau*, vol. 24, no. 3/4, 1933.

Harry O. Wood, Note on the Long Beach Earthquake, *Science*, vol. 78, no. 2022, 1933.

B. Gutenberg and C. F. Richter, Advantages of using Geocentric Latitude in Calculating Distances, *Gerlands Beitr. z. Geophysik*, vol. 40, 1933.

B. Gutenberg, The Propagation of the Longitudinal Waves Produced by the Long Beach Earthquake, *Gerlands Beitr. z. Geophysik*, vol. 41, 1934.

B. Gutenberg, Das "Seismological Laboratory" in Pasadena, *Gerlands Beitr. z. Geophysik*, Supplementband 2, Ergebnisse der Kosmischen Physik, 1933.

B. Gutenberg and C. F. Richter, On P'P' and Related Waves, *Gerlands Beitr. z. Geophysik*, vol. 41, 1934.

B. Gutenberg and C. F. Richter, Contribution to the Study of Deep-Focus Earthquakes, *Gerlands Beitr. z. Geophysik*, vol. 41, 1934.

At a meeting of the Seismological Society of America held on April 12 and 13, 1934, at Stanford University, Dr. Richter presented a paper entitled "Further Results with the Magnitude Scale." Dr. Benioff presented a paper entitled "Instrumental Method for the Determination of the Extent of Faulting." The following two papers were presented by Dr. Richter and Dr. Gutenberg jointly: "Observed P at about 20 Degrees and the Question of Sub-crustal Discontinuities"; "Effect of Crustal Structure on the Reflection of PP."

Harry O. Wood, Maxwell W. Allen, and N. H. Heck, Destructive and Near-destructive Earthquakes in California and Western Nevada, 1769–1933, Special Pub. No. 191, U. S. Dept. Commerce, Coast and Geodetic Survey, 1934.

Chapters by Harry O. Wood in Physics of the Earth, vol. 6, Seismology, Bull. Nat. Research Council, No. 90, October, 1933: chap. 3, pp. 9–31, Volcanic Earthquakes; chap. 6, pp. 41–66, Earthquake Investigation in the Field; chap. 7, pp. 67–82, "Apparent" Intensity and Surface Geology.

Harry O. Wood, Seismological Research in Southern California, *Proc. Fifth Pacific Sci. Cong.*, vol. 3, pp. 2347–2353, 1934.

Hugo Benioff, A New Electro-magnetic Seismograph, *Proc. Fifth Pacific Sci. Cong.*, vol. 3, pp. 2443–2450, 1934.

B. Gutenberg and C. F. Richter, On Seismic Waves (first paper), *Gerlands Beitr. z. Geophysik*, vol. 43, pp. 56–133, 1934.

Hugo Benioff, The Physical Evaluation of Seismic Destructiveness, *Bull. Seismol. Soc. Amer.*, vol. 24, no. 4, pp. 398–403, October 1934.

Halley Wolfe, A Seismographic Recorder, *R. S. I.*, vol. 5, pp. 359–361, October 1934.

Charles F. Richter, An Instrumental Earthquake Magnitude Scale, *Bull. Seismol. Soc. Amer.*, vol. 25, no. 1, pp. 1–32, January 1935.

Harry O. Wood, Earthquakes in California, *Scientific Monthly*, vol. 39, no. 4, pp. 323–344, October 1934. Also issued as: Earthquake Study in Southern California, Carnegie Inst. Wash. Supp. Pub. No. 12, February 15, 1935.

Beno Gutenberg, Crustal Deformations of Gradual Type, *Proc. Fifth Pacific Sci. Cong.*, vol. 3, pp. 1297–1304, 1934.

Beno Gutenberg, The Structure of the Earth's Crust as Indicated by Seismological Data, *Proc. Fifth Pacific Sci. Cong.*, vol. 3, pp. 2511–2521, 1934.

John P. Buwalda and Beno Gutenberg, Investigation of Overthrust Faults by Seismic Methods, *Science*, vol. 81, no. 2103, pp. 384–386, April 19, 1935.

B. Gutenberg and C. F. Richter, On Seismic Waves (second paper), *Gerlands Beitr. z. Geophysik*, vol. 45, pp. 280–360, 1935.

B. Gutenberg and C. F. Richter, On Seismic Waves (third paper), *Gerlands Beitr. z. Geophysik*, vol. 47, pp. 73–131, 1936.

B. Gutenberg and C. F. Richter, Revised and Additional Geocentric Coordinates of Seismological Stations, *Gerlands Beitr. z. Geophysik*, vol. 46, pp. 198–201, 1935.

Harry O. Wood and B. Gutenberg, Earthquake Prediction, *Science*, vol. 82, no. 2123, pp. 219–220, September 6, 1935.

B. Gutenberg and C. F. Richter, Magnitude and Energy of Earthquakes, *Science*, vol. 83, no. 2147, pp. 183–185, February 21, 1936.

Hugo Benioff, A Linear Strain Seismograph, *Bull. Seismol. Soc. Amer.*, vol. 25, no. 4, pp. 283–309, October 1935.

In addition, the annual reports of the Advisory Committee in Seismology from 1921 to 1936 inclusive, printed in Year Books Nos. 20 to 35 of the Carnegie Institution of Washington, present in considerable detail the various approaches to this great problem.

THE SIGNIFICANCE OF PRESSURE AND OF VOLUME IN GEOPHYSICAL INVESTIGATIONS

L. H. ADAMS

Director, Geophysical Laboratory
The Carnegie Institution of Washington

Introduction

In its broadest sense geophysics means the physics of the Earth as a whole, but the term is commonly applied more specifically to the study of the Earth's substance from the standpoint of physics and physical chemistry, and to the determination of the properties and behavior of the rocks and other materials that make up the visible crust. Although geophysics, as sometimes narrowly restricted, is used to describe the practical application of physical methods to the discovery of mineral deposits concealed below the surface, it is important to recognize this as merely one phase of the larger field of the science.

The main features of the Earth, or rather of that part which can be perceived directly, are well known. About one-fourth of its surface is occupied by land, the remainder by water. The uppermost layer of the land generally consists of a thin film of soil covering a thicker layer of sedimentary rocks, such as limestone, sandstone, or shale. These sedimentary rocks, although of wide extent and of great thickness in some places, should be considered as merely a veneer over the underlying igneous rocks. The Earth's crust is now commonly believed to have originated by the solidification of a formerly liquid magma, producing igneous rocks of a great variety of compositions and textures such as the familiar granite and basalt as well as others with different combinations of the rock-forming minerals. Although the chemist by careful analysis has determined the compositions of the various types of rocks, and although the geologist by patient study has learned much concerning the arrangement and past history of rocks, the exact manner of their formation from a molten silicate mass has remained a problem for special experimentation. This great problem has been successfully attacked by an extension of well-known methods the essence of which is the definition and quantitative determination of various simple factors.

Pressure

An investigation of the condition and behavior of an aggregation of materials resolves itself into the evaluation and interrelation of three factors, namely, pressure, temperature, and composition. This is especially evident in connection with the highly specialized researches on systems (or aggregates) containing two or more components. In order to represent the results of such researches, it is customary to construct diagrams in which any two of the abovementioned factors have been chosen as coordinates. In the language of physical chemistry the three variables of equal rank, pressure, temperature, and composition, are the intensity factors of the system.

Of the three factors, pressure is the one with which we are least familiar. We daily encounter articles having a wide range of chemical composition, and we often have our attention directed to very cold objects, such as liquid air, or to very hot ones, such as flames or electric arcs. But in the case of pressure our experience is ordinarily quite limited. The pressure in the water pipes of a house is usually not more than 60 pounds per square inch, or 4 atmospheres, which is a very small fraction of the pressure that is now experimentally realizable and is of geologic importance.

It is essential to distinguish clearly between two varieties of pressure, uniform and non-uniform. The first, often called hydrostatic pressure, signifies a pressure that is the same in all directions. Non-uniform pressure refers to an unbalanced system of stresses, the simplest example of which is a simple thrust or tension exerted in one direction. The latter variety is of great importance in geologic processes involving deformation and flowage, but offers experimental difficulties greater even than those associated with high hydrostatic pressure. For this reason, and also because the phenomena associated with non-uniform or directed pressure are complicated and their interpretation difficult, less progress has been made on the experimental side of this subject than on the purely theoretical.

On the other hand, the technique for the production, measurement, and application of high hydrostatic pressure is well advanced. Given the necessary equipment, the subjecting of material to accurately measured hydrostatic pressures of 12,000 to 15,000 atmospheres can be made a routine operation, and pressures as high as 50,000 atmospheres can easily be obtained. If we recall the fact that in the Earth the pressure increases approximately 500 atmospheres for each mile of depth below the surface, attaining a magnitude of 15,000 atmospheres at the bottom of what is commonly called the crust of the Earth, we see that the experimental pressure range is comparable to that involved in the formation of the Earth's crust, although the pressure in the far interior (more than 3,000,000 atmospheres) is beyond the present experimental range.

Pressure, either uniform or non-uniform, is connected with geophysical problems in two separate ways. The indirect effects of pressure are perhaps the more immediately applicable. Among these is the effect of pressure on the absorption of gases—carbon dioxide and water vapor, for example—by a silicate magma. A pressure of only a few hundred atmospheres will permit a magma to hold a large quantity of these volatile constituents in solution and will influence the manner of crystallization of the rock that is produced. The later stages of the cooling of such aqueo-igneous melts approach closely the conditions of pegmatite and vein formation.

The direct effects of pressure are of many kinds. Nearly all physical properties are changed to a notable extent by the application of pressure. Of especial interest is the effect on the melting point and solubility of rock-forming minerals, and also the probability that new and denser forms of various minerals may be produced under the conditions of high pressure existing at even moderate depths within the Earth. The effect of pressure on volume will be treated separately in a following section.

These considerations should not obscure the fact that pressure is of importance only in so far as its relations to other variables are understood. The significance of pressure lies in its allocation as one of the three principal intensity factors—pressure, temperature, and composition.

Specific Volume and Density

For purposes of the present discussions volume may be thought of as one of the fundamental quantities dependent on the above-mentioned factors. The volume occupied by a unit amount of material is called the specific volume; its reciprocal is the density. Many measurements on specific volume or density have been published and it might seem that this field has not been neglected. There are a number of books devoted solely to tables of density data, and in the International Critical Tables, which contain a condensed and selected summary of all the data on the subject, it was found necessary to use about three hundred large-sized pages for the presentation of this material. And yet a great part of the data is of little use, principally because of insufficient accuracy. It is indeed very strange that this should be so, for the measurement of density is one of the simplest of determinations. With a relatively crude apparatus density can be measured to six significant figures, and density differences to the eighth or ninth figure after the decimal. In spite of this, the densities of NH_4NO_3 solutions, as given in the International Critical Tables on the basis of the best available data, were found in the course of a recent investigation to be in error by as much as 8 units in the fourth decimal place, and had to be redetermined before the main investigation could proceed.

The generally poor quality of the recorded measurements of densities, especially those of solutions, is a continuous handicap to investigators in many fields. It would be quite feasible for one person, of average training, equipped only with a balance and some glassware to obtain in a reasonable period of time more really useful data on the densities of solutions than the total now available. Probably the reason that there have been so few serious efforts to make accurate determinations is that the careful measurement of density has seemed a tiresome and profitless task. In the absence of a definite objective it would of course be difficult to become enthusiastic over such an undertaking. There is involved also the common urge for the complex and elaborate in apparatus and experimentation. Certain kinds of investigation, it is true, require complicated machinery and numerous tools and appliances, but on the other hand accurate measurements as well as great discoveries have often been made with simple apparatus. It may not be necessary to limit ourselves to the primitive instruments of the earliest investigators, or to be like Aristotle, who was compelled "to fix time without a watch, to compare degrees of heat without a thermometer, to observe the heavens without a telescope, and the weather without a barometer," and who "of all our mathematical, optical and physical instruments possessed only the rule and compass, together with the most imperfect substitutes for some few others." Some varieties of measurement require elaborate apparatus, but other very useful ones do not. Not many years ago the element argon was discovered in the atmosphere because attention had been drawn to the apparent discrepancy between the densities of atmospheric nitrogen and synthetic nitrogen. Although the density difference is readily measurable it had escaped detection. The importance of density measurements in connection with X-ray studies is not generally appreciated. Wave lengths of these radiations depend ultimately on the spacing in some standard crystal. This spacing is determined simply by measuring the density of a suitable standard, for example calcite or common salt.

Another important application of density measurements is to the analysis of solutions. The composition of a mixture may sometimes be determined with remarkable precision by measuring its density. This method is especially convenient for analyzing liquid solutions, but it has not received the attention that it deserves. For aqueous solutions of ordinary salts it is easy to obtain an accuracy of better than 0.001 per cent.

A knowledge of density is especially useful in the study of minerals. In many instances the density serves as an important means of identification. If the mineral belongs to an isomorphous series, such as the plagioclase series (ranging from albite to anorthite), or the olivine series (with varying amounts of iron replacing magnesium), a density measurement may afford an important check on the optical determinations. Furthermore, if an extraneous mineral is present, a combination of the two sets of measurements will provide an accurate estimate of the amount of the other mineral. The available density data for minerals and rocks, although very numerous, are generally inferior in accuracy. If better data were at hand the density could often serve as a valuable analytical aid. Although extreme accuracy may not be justified for some minerals because of impurities, or in the case of many rocks because of porosity, an accuracy somewhere between one unit in the first decimal place (which is often encountered) and one unit in the fifth decimal place (which can be obtained without difficulty) would constitute a reasonable and usable precision.

Volume Changes

So much for the volume occupied by materials under ordinary conditions and for some of the aspects of this simple, primitive quantity. Let us now consider briefly the phenomena accompanying changes in volume. The most familiar means by which the volume of a substance may be altered is by varying the temperature. Volume changes may be gradual as, for example, when the material remains homogeneous, or discontinuous as in the case of the freezing of a liquid or the condensation of a gas. It is an interesting fact that only with great difficulty can the volume of liquids or solids be changed significantly. More especially, decreases in volume much below the usual volume are limited in extent. If a typical solid or liquid is cooled, without change of state, down practically to absolute zero, it will ordinarily shrink in volume by less than 2 per cent. It appears that solids and liquids stubbornly resist change; in sharp contrast with gases, they seem very reluctant to allow their characteristic volume to be greatly altered.

The most effective means for diminishing volume is the application of high pressures, that is to say, pressures measured in thousands of atmospheres. We may thereby decrease the volume of typical liquids to about three-quarters of their original volume. Solids are usually more resistant, but it is not unlikely that under pressures considerably higher than those possible to obtain in the laboratory, their volume could be made to shrink by an amount similar to that actually observed for liquids. Both solids and liquids become steadily more difficult to compress as the pressure is increased. For reasons given below, the application of a pressure of a million atmospheres would not reduce the volume of ordinary liquids and solids by a very striking amount.

The facilities and technique for this field of investigation are a comparatively recent development. A little more than half a century ago the real foundation of this branch of science was laid by Amagat. Somewhat later, further advances were made by Tammann, and many investigations on the behavior of substances under pressures up to 3000 atmospheres were carried out. In the present century a more striking development was made by Bridgman, who extended the pressure range far beyond what had previously seemed possible and has made many observations on the effect of decreasing the volume of elements and compounds. Although experimentation in this field has been confined to a very few laboratories, and although the maximum pressure at which useful measurements can now be made—about 50,000 atmospheres—is still insufficient to apply directly to some geophysical problems, the information already collected is sufficient to show that the behavior of materials under diminished bulk is a fascinating and instructive subject. Physical properties are altered in curious ways, new forms of matter make their appearance, chemical compounds undergo unexpected reactions, albuminous materials are coagulated, and bacteria and other organisms are destroyed.

The relative amount by which the volume is diminished at a given pressure is called the compression, and the rate of change with respect to pressure is called the compressibility. The measurement of the compression of solids with satisfactory accuracy offers some difficulties, but it has been found possible to carry out such measurements successfully on a great variety of materials including rocks and minerals. Results such as these have been applied to the solution of many interesting problems, of which two will be mentioned briefly.

It had long been suspected that the interior of the Earth consisted of metallic iron, without any very convincing reason being given to justify the supposition. The laboratory measurements on the volume change of rocks under pressure supplied the first definite evidence on this question. As is known from the size and mass of the Earth, its density, $5\frac{1}{2}$, is about twice that of the usual rocks found at the surface. From this it follows that at the center the density probably is at least 10, if the average density is to have the correct value. There are two possible ways by which the high density in the central portion of the Earth may be explained. First, under the enormous pressures in the interior of the Earth silicate rocks might be squeezed into a bulk about one-quarter as small as the original amount, thus allowing the interior to have the requisite high density. The second possibility is that in the central portion is a core of intrinsically heavy material.

A choice between these alternatives, and a partial solution therefore of one of the great problems of geophysics, was afforded by the utilization of measurements of the volume changes of rocks under pressure, in combination with other pertinent data. It turns out that the volume of a solid silicate cannot be reduced to one-half its original volume even by the application of the enormous pressures in the Earth's interior. Because considerably more reduction than this would be required in order to explain the high central density on the basis of the first of the above-mentioned alternatives, it was concluded that the Earth has a central core of a dense substance, which in accordance with other considerations is judged to be metallic iron.

Measurements of volume have been used also to determine the nature of the deep-seated material outside the core. From the compressibility of a rock it is possible to calculate the velocity with which elastic waves, such as those which accompany sound or other vibrations, will travel through the material. Furthermore, seismologic data furnish us precise values for the velocities of earthquake waves at various points of the curved paths followed by these vibrations in passing from the earthquake center to the recording instrument. The data can be readily expressed in terms of the wave velocity at various depths below the surface. Therefore, by comparing these with the velocities obtained from the laboratory measurements on the compressibilities of various types of rock, we are able to draw important conclusions concerning the composition of the material at various depths within the Earth. It has been found, for example, in agreement with inferences drawn from geologic observations, that the upper layer of the Earth's crust is predominantly granitic and that below the crust we have a highly elastic material, presumably similar to the rare rock known as peridotite.

Pressure, Volume, and Equilibrium

Equilibrium, in physical chemistry, refers to the mutual stability of various solid, liquid, and gaseous phases that may be present in a mixture of one or more components. A complete study of a system, whether it be a simple mixture of salt and water, or a combination of several rock-forming oxides, requires a determination of the behavior of the system under all possible conditions of temperature and pressure. Of the many investigations of physical-chemical equilibria, nearly all have been restricted to a pressure of one atmosphere. A large amount of data has been secured concerning the fusion temperatures of mixtures and concerning the effect of temperature on solubility, but very little information is available as to the corresponding effect of pressure. The study of systems in which the volume has been considerably reduced by the application of high pressure enlarges our view and is quite essential for an understanding of the processes by which the Earth attained its present condition. Expressed in conventional language, what is needed is a set of diagrams representing freezing points or solubilities at various pressures.

Volume has a special significance in the determination of equilibrium under high pressure. In a solution the quantity that is most useful is not the volume of the separate constituents, but a derived quantity, sometimes called the partial volume. It is well known that the equilibrium between a solid and a liquid, or any set of phases, is best described in terms of a readily measurable quantity called the chemical potential. It happens that this equilibrium indicator changes in a very simple way with pressure. Indeed, at a given temperature the change in chemical potential is simply the product of the pressure change and the volume or partial volume. Hence, in order to determine the behavior of a system under high pressure, we measure, at a series of pressures, the volumes of the various solutions and other phases that may be present, and thence pass by a straightforward calculation to the solubility at a given pressure or the freezing pressure at a given concentration. The conditions under which a transformation occurs are thus determined without causing the transformation actually to take place; for the direct measurement of a freezing point we substitute an indirect determination based principally on the measurement of volume. This indirect method has proved to be a valuable aid in the study of systems under pressure and has already been successfully applied to several binary systems. It appears to offer the most promising line of attack on the problem of the melting behavior of rocks at considerable depths below the surface of the Earth.

Volume, A Fundamental Concept

If we direct our attention to the progress of science throughout the ages, we may note two outstanding characteristics. The first is the development of science at an ever increasing rate, and the second is a recurrent tendency toward simplification. On the one hand we have the steady accumulation of a vast array of facts in every branch of science, and on the other hand we observe the patient efforts of investigators to classify and arrange the separate items of information that have been collected, and to express them finally in simple terms. Especially in physics and chemistry, investigators have continued their efforts to describe all the phenomena of Nature on the basis of a limited number of fundamental concepts combined with a few great generalizations.

In the early days of philosophy the four elements, fire, air, earth, and water, were considered to be the basic concepts by which all natural phenomena could be explained. More recently the notions of force, energy, and heat were combined and given a prominent position in the generalizations of physics, and finally, by the revolutionary advances of Einstein and others, space and time were welded into one homogeneous entity in which all physical attributes, even mass and gravitation, have been submerged, to reappear as mere irregularities on a four-dimensional space-time framework. For purposes of the present discussion the important fact is that the process of simplification has been carried to an extreme limit, the number of fundamental concepts being reduced to one, namely, space-time. Thus all phenomena, however complex, may be described in terms of a single entity. In the physical world generalized space has therefore been given a place of leading importance.

Admittedly these considerations are less pertinent to some other branches of science. Biology is concerned more especially with the qualities of living organisms than with the laws that govern their motion; mathematics and philosophy, unhampered by experiment, devise their own concepts to suit the purpose in hand. Nevertheless, from the standpoint of the properties of matter it is useful and suggestive to consider space as the basic entity, the fundamental nature of which readily becomes evident to us even without the abstract reasoning of the mathematical physicist. Our first perceptions are those of the form and position of objects in surrounding space, and our daily activities involve to a considerable extent the notions of extension, or size, or volume. For us, as three-dimensional beings, it is of course natural that in practical matters we should direct our attention mainly to three-dimensional space rather than to the more complete and elegant space-time.

Concluding Remarks

The conclusion to be drawn from these considerations is that space, extension, or volume, is the oldest, the simplest, and the most fundamental of all the physical attributes. In experimental science volume is a quantity that occupies a unique position. Pressure stands out as the agent by which volume can be most effectively altered. In geophysical investigations pressure enters as an intensity factor, coordinate with temperature and composition, and of equal importance. A knowledge of the effect of pressure on volume, together with more complete data on the volume occupied by materials under ordinary conditions, is of great consequence. The intrinsic interest and wide applicability of information on these elementary quantities deserve more general recognition.

WATER IN GEOLOGICAL PROCESSES

GEORGE W. MOREY
Geophysical Laboratory
The Carnegie Institution of Washington

Water as a factor in geological processes has long been a subject of speculation, speculation which has been guided only in part by pertinent theory, and has rarely been controlled by experimental demonstration. Water has left its mark on all stages of the geological process, but the laboratory approximation to natural processes is made difficult by the necessary high temperatures and pressures, and only in recent years has an adequate experimental technique been developed.

Occurrence of Water in Magmas

That water is an original constituent of the molten mass or magma from which, on cooling, the igneous rocks have crystallized is generally accepted. Direct evidence is furnished by the volcanic eruptions, the gases from which are largely water. Enormous quantities of water are poured into the air in volcanic eruptions, and much of the destruction caused by catastrophic eruptions such as those of Mount Pelée and Mount Lassen is attributable to a blast of superheated steam. Other evidence is inferential, and some of it is given later, but the occurrence of hydrated minerals such as the micas in deep-seated rocks is pertinent.

The importance of water has been realized for many years. Bunsen, in 1851, stressed the importance of water in rock-forming processes. Writers of the French school, notably St. Claire Deville and Michel-Lévy, emphasized especially the influence of *fluides minéralisateurs* in the differentiation of a magma into many rock types. Later writers have dealt more with the mechanism of the action of water, and the physical condition, liquid or gaseous, of the end stages of magmatic differentiation.

Concerning the quantity of water present at different stages of the process of differentiation nothing is known, and until experimental data bearing on the problem can be obtained the problem will remain in the realm of speculation. Probably most geologists believe that the early or basaltic magma contains much less water than the later or granitic stages, but some question even this conclusion. Estimates of the actual amount of water in a granitic magma range from 1 to 8 per cent by weight or even more, but until estimates can be checked by experiment they remain unsupported guesses.

Accumulation of Water in Magmas

The primary magma is assumed to be low in both silica and water, and the trend of rock formation is to separate first anhydrous minerals even lower in silica than the parent magma. The resulting residual magma becomes richer both in silica and in water, and the relationship between these two components in affecting the late stages of rock formation is of paramount importance. The increase in silica content leads to the formation of granite as one of the last products of igneous action. Granite is composed chiefly of highly siliceous acidic feldspars, of mica, a mineral containing water,[1] and of silica as quartz. When the changing magma, changing by the crystallization of less silicic or basic minerals, such as olivines, pyroxenes, amphiboles, and plagioclase feldspars, has progressed to the stage of granite formation, the water content may have increased in amount until its influence has become dominant.

That the granites, and especially those coarse-grained aggregates of much the same minerals which are called pegmatites, are formed at comparatively low temperatures is shown by the application of that valuable "geologic thermometer," quartz. Quartz exists in two different crystalline forms, one deposited above the "transition temperature," 573°; the other, below that temperature. When the high-temperature form is cooled it invariably changes to the low-temperature form, and the fact of that change can be deduced from a study of the quartz under the microscope. Application of this geologic thermometer has shown the quartz from some vein and pegmatite deposits to have been deposited below 573°. Granites and pegmatites represent late stages in the differentiation of a molten magma.

---

[1] The feldspars of granite are: predominatingly orthoclase, $K_2O \cdot Al_2O_3 \cdot 6SiO_2$; albite, $Na_2O \cdot Al_2O_3 \cdot 6SiO_2$; and anorthite, $CaO \cdot Al_2O_3 \cdot 2SiO_2$. The micas are a complex group of hydrated minerals, of which muscovite, $K_2O \cdot 3Al_2O_3 \cdot 6SiO_2 \cdot 2H_2O$, is perhaps the simplest.

Effects of Water in Magmas

The most important effects of water in the formation of deep-seated igneous rocks are to lower the freezing point of the magma and to lower its viscosity greatly. The lowering of the freezing point is a necessary thermodynamic consequence of increased H_2O content; an additional component almost invariably[2] results in lowering of freezing point, and water is particularly efficaceous by reason of its low freezing point and comparatively small molecular weight. Granites are late stages in the sequence of formation of igneous rocks from molten magmas, yet the characteristic minerals of granite, quartz, and feldspar have high melting points. The presence of water lowers the melting range, and is the direct cause of the low temperature of formation of these important rocks, the commonest igneous rocks of the Earth's crust.

The great effect of water in reducing viscosity may be inferred from observations on granites and pegmatites. As mentioned above, they are the products of late stages in the process of differentiation of the magma into various mineral species. The essential minerals of granites, the highly siliceous feldspars, and silica itself, stand out among all other compounds by reason of the extreme viscosity of their anhydrous melts. The viscosity of these feldspars at their melting points is about ten thousand million times greater than that of water at its melting point, and the viscosity of the pure minerals at the temperature of pegmatite formation would have increased to an even more astronomical magnitude.[3] Their enormous viscosity is the reason why these feldspars have never been crystallized in the laboratory without the aid of mineralizers; yet crystals several feet in length are often found in the low-temperature pegmatites. Some agency must have profoundly affected the viscosity of these melts, and that agency was water. The influence of water in lowering melting point and decreasing viscosity plays a dominant part in the formation of intrusive rocks.

---

[2] The exceptions to the general rule probably are not met with in magmatic systems.

[3] The viscosity of water at 0° C is 0.018 poises; of castor oil at 20° C, 9.86 poises; of albite at its congruent melting point, about 10^7 poises; and of orthoclase at its incongruent melting point, about 10^8 poises.

The greater fluidity resulting from the presence of water in the magma makes possible the differentiation of the magma into many rock types. An important process by which this differentiation is accomplished is in the settling or floating of crystals in the liquid under the influence of gravity. Were no water present the viscosity of the anhydrous magma could greatly prolong this process or, with the more viscous magmas, prevent it altogether. Another important mechanism of differentiation is the squeezing-out of the liquid from a mush of crystals and liquid, a process greatly facilitated by the decreased viscosity resulting from the presence of water. In other cases the liquid magma, or even a mush of crystals in magma, is pressed into thin veins and dikes, and the abundance of these intrusions probably is a consequence of the fluidity resulting from the presence of water.

Water in the magma differs from the usual rock-forming constituents in that at magmatic temperatures pure water is a gas. Indeed, for many years it was not understood how water could be in a liquid mass at temperatures above a red heat, far above the critical temperature of water, above which pure liquid water cannot exist no matter how great the pressure. But the water is not pure; it is one constituent of a complex liquid, and in such a solution the critical temperature of water is without significance. At the temperatures in question, pure water would be a gas, just as at temperatures above the critical temperature of carbon dioxide, 31.1° C, pure carbon dioxide would be a gas; but, just as carbon dioxide can remain dissolved in liquid water above 31.1° C, so water can remain dissolved in a liquid magma above 374°.

Because of its volatility, water differs from the non-volatile constituents in that it requires pressure to hold it on the job. As the magma cools, its water is concentrated in the residual magma and the system as a whole is unusual in that the pressure increases as the temperature falls. So long as crystallization does not take place, and the liquid remains of constant composition, pressure falls with temperature as usual, but when the anhydrous silicate minerals are removed from the liquid by crystallization, the effect of the increasing water content at first greatly overshadows that of the falling temperature. Indeed, this effect gives us an important clue to the causes for some outstanding features of volcanic action.

Water in Volcanic Eruptions

As cooling and consequent crystallization continue, the pressure developed by the water accumulating in the liquid magma rises, and the subsequent history depends on the strength of the restraining crust. The usual case is that the crust is competent; the process of magmatic differentiation continues without interruption through the stage of pegmatite formation down to the later stages of fumarolic and hot spring action, in which pressure falls with temperature. But if the crust is not competent, the building up of pressure continues until relieved by rupture of the crust, and a volcanic eruption ensues.

A volcanic eruption is an accident which sometimes appears in the course of the usually orderly process of magmatic differentiation, and the type and severity of the eruption depend on the physical condition of the overlying material. If the crystallizing chamber is surmounted by a vent in which there is liquid lava, that lava will act as a safety valve. When the pressure has built up to equal the weight of the overlying column of lava, the excess pressure will escape without difficulty, and there will be a continuous and comparatively quiescent volcanic eruption. The mild bubbling of Kilauea is of that type, and the rate of crystallization in the chamber below determines the intensity of the more or less continuous activity observed at the surface.

When a greater pressure is required to effect the escape of the imprisoned water, the eruption is deferred to a later stage in the differentiation. By this time, the water content has greatly increased, as the result of the removal by crystallization of a large part of the basic minerals. The non-volatile constituents correspond to the more siliceous minerals of granite. These are inherently of high viscosity. Because of the high pressure, many of the resulting eruptions are of the catastrophic type; because of the high viscosity of the water-free material, some of the ejecta are highly vesiculated. Examples are the famous eruption of Krakatau, the dust cloud from which took two years to settle, during which time it made several circuits of the Earth in the upper atmosphere, and piles of fine pumice five feet thick covered the sea twenty miles away. The eruption of Mount Pelée was of this type, and the famous and terrible "nuée ardente" that destroyed the city of

St. Pierre, Martinique, with the loss of thirty thousand lives was a blast of dust-laden superheated steam.

Water and Magmatic Ore Deposits

Many types of ore deposits are associated with intrusive rocks, and in their formation water is believed to be an important factor. Possibly most ore deposits are formed at late stages of magmatic differentiation, but some primary magmatic segregations take place in the earliest stages. In such segregations the effect of water is chiefly in lowering the melting point of the magma. Other types of deposits owe their existence to aqueous emanations or exhalations from the magma, which penetrate the country rock, sometimes for great distances, and alter its composition both by deposition of material from the magma and by removal and redeposition of the parts of the invaded rock. The exact mechanism of this action is still uncertain, and much experimental work will be necessary before the complicated process is clearly understood. It probably is associated with the ability of gaseous emanations to traverse pores and fissures too small for the liquid magma to penetrate, and with a real "solubility" of the ore minerals in the superheated and highly compressed water vapor. Many deposits of valuable minerals are among the final products of differentiation, representing accumulations of originally minor constituents, concentrated in the residual aqueous solutions by the progressive deposition of silicate minerals until at comparatively low temperatures the ore-bearing solutions contain silicates only as minor constituents.

Water and Sedimentary Rocks and Ores

The influence of water is not confined to the rocks and minerals that are formed by fusion. Over much of the Earth's surface the igneous rocks are covered by a relatively thin layer of sedimentary rocks, which in places have been changed by the action of pressure, heat, and water into metamorphic rocks resembling the igneous rocks. The sedimentary rocks are formed by the decomposition of the primary igneous rocks by water. All silicate minerals are decomposed by water at ordinary temperatures, at rates dependent on the composition of the mineral, its physical condition, and the temperature. The process is not one of pure solution, but rather one of decomposition, in which certain constituents, potash or soda, less rapidly lime and magnesia, are leached out, leaving iron, alumina, and silica to form clays and sand.

The dissolved and eroded materials are in part carried away by the water, some in solution, others carried in suspension as fine particles; that remaining behind forms deposits of such minerals as clay. The dissolved and transported materials in large part reappear as secondary minerals. Lime is deposited as carbonate, later compacted by water and pressure into limestones; sometimes, with the aid of water, pressure, and heat, metamorphosed into marble. The sand is deposited in beds, later cemented into sandstones by water-borne calcium carbonate or silica, sometimes metamorphosed with the aid of water, heat, and pressure into quartzites. The sedimentary rocks result from the decomposition of the igneous by water, the transport of the material by water, and the subsequent consolidation by water; the metamorphic rocks are a further stage in the same process, in which the action of the water is magnified by heat.

Many of the ore deposits are products of this stage of aqueous activity. The activity of water in the upper layers of the Earth's crust, the zone of oxidation, is one of decomposition and solution; in the lower zones these materials are redeposited. This mechanism, by which valuable substances widely disseminated in igneous, sedimentary, or metamorphic rocks are transported and redeposited in workable quantities, is of frequent occurrence. Placer deposits result from the mechanical sorting by water of products of rock decomposition, and some types of mineral deposits, for example, the oölitic iron ores, are the result of chemical reactions in the surface waters.

Laboratory Study of Water in Magmas

The experimental study of the action of water is made difficult by the high pressures and temperatures that are required. Much of the work of the Geophysical Laboratory has been on the phase equilibrium relationships of non-volatile rock-forming constituents, which required the development of a technique for experimentation at high temperatures, but at ordinary pressure. The vapor phase was absent or could be ignored. Most work at high pressures has been with purely hydrostatic pressures, in which pressure is exerted only by or through a liquid. Also, most of the work has been at ordinary or slightly elevated temperatures. But to solve the problems involving water requires, as the first step, knowledge of phase equilibrium relations of silicate systems containing water with a vapor phase present, at high temperatures, and under the resulting high pressures, a circumstance which greatly increases the experimental difficulties and requires the development of new techniques and new forms of apparatus.

The problem has been approached by many investigators, and an extensive literature dealing with hydrothermal formation of silicates exists. A recent bibliography[4] listed 156 papers, the first of which appeared in 1845, and 67 of which were published before 1900. The experiments, as a rule, were made by heating materials with water in closed bombs, under conditions which were not carefully controlled, and under which equilibrium could not be attained.

Such studies give no information as to the solubility or fusion relationships in the aqueous systems. The time is past in which the mere artificial production of a mineral, even by hydrothermal methods, is of more than passing interest, unless that mineral in itself is exceptional. Systematic information is necessary, which primarily means knowledge of the phase equilibrium relationships in systems containing both water and silicates. Aqueoigneous fusions under controlled conditions of temperature and pressure, similar to those dry fusions the study of which in the Geophysical Laboratory has contributed so much to scientific petrology, are necessary for further progress. Merely heating silicates with water in closed bombs, and obtaining thereby crystallization of minerals under conditions bearing no known relationship to the conditions of formation in nature, adds little to our knowledge. Quantitative information as to the interrelationship of every variable is essential.

The problem of the effect of water in mineral formation has long been a part of the program of the Geophysical Laboratory. In the discussions and reports leading to the establishing of the Laboratory, the study of phase equilibrium relations in systems containing water was stressed by a distinguished group of geologists, and Van Hise, in his "Report on Geophysics" in volume 2 of the *Carnegie Institution of Washington Year Book* (page 180) said:

> Experiments should be carried on with aqueous solutions under various pressures and at various temperatures. The higher temperatures should approach

---

[4] George W. Morey and Earl Ingerson, "The Pneumatolytic and Hydrothermal Alteration and Synthesis of Silicates," *Econ. Geol.*, vol. 32, pp. 607–761, 1937.

those of magmas, in order that the relations of crystallization from magmas and crystallization from water may be learned.

The first approach to the problem was by the conventional method of heating the non-volatile components with water in steel bombs, but the inadequacy of that method was soon realized. A fortunate discovery that an aqueous melt of potassium silicate could be chilled from high temperature without change in the composition or condition of the melt by plunging the steel bomb into water led to the first quantitative study of a water-silicate equilibrium. For example, the melting point of potassium disilicate could be lowered from 1044° C to 500° C by heating in a steel bomb with the requisite amount of water. The resulting melt, after freezing or quenching in water, was a hard, brittle glass containing a little less than 10 per cent of water by weight. At the temperature of the experiment it was a fluid melt, the water in which exerted a vapor pressure of 66 atmospheres. By repetition with different mixtures of potash and silica, at different temperatures and pressures, the entire system was worked out, and it was demonstrated that aqueous silicate solutions could be studied at temperatures above the critical temperature of water.

This was a special adaptation of the "quenching" method, developed in the Geophysical Laboratory for the study of silicate equilibria. Its development was a matter of necessity. The outstanding characteristic of silicate melts in general is the sluggishness with which they crystallize. This is the property which makes the glass industry possible. Because of their sluggish reactivity, the ordinary methods of melting point determination can be applied only with difficulty or not at all. But this same sluggishness makes it possible to remove a melt from a furnace and cool it quickly enough to "freeze" the equilibrium, to obtain the charge as a glass the condition of which with respect to crystallization is identical with that of the melt at the temperature of the experiment. The quenched sample is then examined with the microscope. If no crystals are present, the temperature of heat treatment was above the melting point; if crystals are present, the temperature was below the melting point; and by decreasing the temperature interval between experiments, the melting point can be located as closely as desired.

An apparatus based on a similar principle has been constructed for studying the effect of water. It consists of a stainless steel bomb, heated by an enclosing electric furnace and capable of withstanding the necessary pressures, inside of which is an independently controlled electric furnace equipped with a thermocouple for measuring the temperature of a charge of silicate being studied, and an electrical device for quickly cooling that charge at the end of an experiment. The steam supply is furnished by a separate boiler, also capable of withstanding high pressures. If a silicate is heated above its melting point in a closed vessel containing steam at a given high pressure, the molten silicate will dissolve water, and the dissolved water will lower the melting point. By making a series of heatings at various temperatures and at a constant pressure, the lowering of melting point by the water dissolved at that pressure can be determined. The amount of water dissolved can be determined by analysis, and quantitative values thus obtained for the temperature, pressure, and composition. Repetition at a series of pressures enables the entire system to be worked out, and gives promise of replacing uncontrolled speculation in geological theory with precise experimental data.

An experimental technique which makes possible the study of aqueous solutions at high temperatures, under the high pressure developed by such solutions, will result in the gradual accumulation of knowledge fundamental not only to geological science but also to all branches of chemistry. Knowledge of aqueous solutions is a large part of chemical science, and a large part of chemical technology makes use of aqueous solutions for the transformation of raw material into products useful to man. Present-day chemistry is almost entirely the chemistry of solutions at low temperatures, below the boiling point of water, but the chemistry of solutions at high temperatures and pressures offers untold possibilities for the development of a new technology whose future may surpass that of the technology of today.

THE SURFACE OF THE MOON

OBSERVATIONS BY THE COMMITTEE ON STUDY OF THE SURFACE FEATURES OF THE MOON

F. E. WRIGHT

Geophysical Laboratory
The Carnegie Institution of Washington

The Carnegie Institution of Washington was founded by Andrew Carnegie to advance knowledge in different fields of science through research work. In the early days of the Institution careful surveys of certain branches of sciences were made by special committees of experts to ascertain the particular problems in each field that were strategically important and could be attacked by scientists working together but approaching each problem from different standpoints. The value of diversified but coordinated attack on a problem is well recognized in the military services (tactics) and is equally important in science.

The committees, after study of the different fields, recommended certain problems for consideration. Of these a limited number were adopted and small groups of research men were brought together and given opportunity to work on them. These groups were gradually organized into the departments and laboratories of the Institution. A number of these units have now been operating for more than thirty years, and have made extremely important contributions to science. In addition to support of the work of the departments, the Institution has aided many individual investigators (research associates) on specific problems related to work of the Institution.

During the first twenty years of their existence the departments were so occupied with their own problems that each one functioned as an independent unit with little knowledge of the work and personnel of the other groups. This situation was recognized by Dr. Merriam when he became President of the Institution in 1921. On his initiative, steps were taken to bring the several departments into closer touch with one another and to a better understanding of the purposes and functions of the Institution itself.

At the same time he realized the desirability of more effective contact with the public in order that the work of the Institution might be better known and of greater service. He felt that the public had the right to know of the work of a semipublic institution, like the Carnegie Institution, operating under special congressional charter and devoting its energies entirely to research work in science. To meet this responsibility several steps were taken. At the time of the Trustees' annual meeting greater emphasis was placed on the public exhibition of the results of current work of the Institution; the number of public lectures by members of the Institution was increased; and many of the lectures were published. The annual exhibition and the annual series of Institution lectures have been successful and have attracted wide attention. Under Dr. Merriam's guidance, the Editor of the Institution has prepared a series of bulletins for distribution to newspapers (news releases), to schools and other organizations (news bulletins), and to staff members of the Institution (staff edition of news bulletins). The purpose of these bulletins is to present, in easily readable form, brief statements on some of the investigations by the Institution; serious effort is made to have the bulletin articles correct and authoritative. The large demand by the public for these bulletins is evidence that they are meeting a need and are appreciated.

Dr. Merriam has also encouraged greater cooperation between Institution departments by appointing interdepartmental committees for study of problems in fields peripheral to those of two or more departments. A good example of a committee of this kind is the Committee on Study of the Surface Features of the Moon.

Realizing how little definite information was available on the nature of the materials exposed at the moon's surface and on the mode of their distribution, Dr. Merriam organized the Moon Committee and appointed to it certain members of the staff of Mount Wilson Observatory (Drs. W. S. Adams, F. G. Pease, E. Pettit) and of the Geophysical Laboratory (Drs. A. L. Day and F. E. Wright, Chairman), and from the group of Research Associates (Drs. J. P. Buwalda, P. S. Epstein, and H. N. Russell). The Committee consisted of four astrophysicists, one mathematical physicist, one volcanologist, and two geologists. With the exception of

Dr. Pease, who for many years had been enthusiastic over the possibilities of research work on lunar surface features, none of the members had made special study of the problem. They were given no specific instructions, other than that implied in the name of the Committee, but were afforded opportunity, as time permitted, to survey the problem and, with the unusual instrumental facilities available within the Institution, to investigate any phase of it that might seem promising.

The Committee has now been at work for a number of years gathering definite data of measurement on the nature of the surface materials, on their distribution, and on the characteristics of the surface features themselves. Throughout this period the members have cooperated effectively and given generously of their time.

Early in the attack on the problem it was realized that precise data of measurement are needed rather than suggestions regarding possible modes of formation of the different lunar surface features. Before the origin of these features can be adequately discussed, knowledge of the nature of the materials exposed at the surface of the moon is essential; also information on the conditions that exist at the lunar surface. A photographic map consisting of a series of photographs of the moon at different phases and with the plane of mean libration as its plane of projection is needed; also a topographic map of the central part of the moon's disk. Information is desired regarding the exterior ballistics of projectiles as represented by ejectamenta from volcanoes or from craters produced by impact of large meteorites or of swarms of meteorites. Statistical information on the frequency of lunar craters of various sizes is required.

Work on these and related problems has been in progress for some years and the results obtained have increased our knowledge of lunar surface features. Some of these results we shall now consider briefly.

Nature of Lunar Surface Materials

In study of rocks and geologic formations on the earth the geologist examines their occurrence and relations to other formations in the field, and procures samples of each type of rock for further tests in his laboratory. He prepares thin sections of the rocks and examines them under the petrographic microscope; he makes chemical tests and, if necessary, may make a complete chemical analysis of an important sample. Under the microscope he determines each mineral by the effects it produces on polarized light on transmission through the thin mineral sections. These effects are many; in general, they are easy of measurement, so that quite accurate and satisfactory determinations of the minerals composing a rock are possible.

In the study of lunar surface materials the geologist cannot visit the moon and collect samples; he must be satisfied with measurements at the long-range distance of about 240,000 miles. The rays of the sun after reflection by the moon reach the earth as moonbeams 3 seconds later; they serve as lunar messengers and carry signals in code from its surface. It is our task to receive and decipher these signals if we are to ascertain the kinds of substances that are exposed at its surface.

The moon shines by reflected sunlight. The sun's rays impinging on its surface undergo, on diffuse reflection, changes that can be measured. These changes are of three kinds: (*a*) a large part of the energy of the incident sunlight penetrates a short distance into the surface and is absorbed, thereby heating the surface materials; the heated surface in turn radiates the absorbed energy into space as heat waves (planetary heat); (*b*) different wave lengths or colors of the incident sunlight may be absorbed in different proportions, whereupon their relative intensities are altered slightly, thus coloring the scattered light (selective absorption); (*c*) a certain amount of plane polarization is introduced on scattering and reflection; the amount depends on the character of the material and on its surface. In general, dark-colored opaque substances polarize the light much more than do non-opaque substances into which the incident light may enter and be returned in part by internal reflections.

Detailed observation of these changes and comparison of the results with those obtained on terrestrial materials under similar conditions of illumination enable the observer to ascertain the general nature of lunar surface materials. The problem is analogous to that of the determination of minerals with the aid of the petrographic microscope. In the one case the changes produced in light on diffuse reflection are analyzed; in the second case, those resulting from transmission through mineral sections and grains are determined. With lunar substances, unfortunately, the changes are relatively few and cannot be measured with the degree of accuracy possible in polarization phenomena of light transmitted through mineral plates. In lunar observations it is not feasible to apply the methods of transmitted light; the observer has no choice but to infer the nature of the surface materials of the moon from the slight changes they induce in sun's rays on reflection. If satisfactory conclusions are to be drawn, it is essential that the measurements be made with high accuracy by independent methods and through several complete lunations for each method. It is for this reason primarily that the task is time-consuming and difficult.

The proportion of plane polarization in a beam of moonlight can be ascertained by several methods based on various types of receivers, such as the human eye (visual), the photoelectric cell, the thermoelement, or the photographic plate in a polarization spectrograph. These methods cover different parts of the spectrum, but their spectral ranges overlap sufficiently so that the results obtained by one method can be checked by those obtained by other methods.

Polarization eyepiece. Visual determinations of the percentage amounts of plane polarization in light from different parts of the moon at different phases of the moon have been made by the Moon Committee during eleven lunations. As a result, the changes in amount of polarization with change in lunar phase angle are now known with a fair degree of accuracy for the light received from twenty-four selected areas on the moon's surface and also for the integrated light from the entire visible lunar surface. Study has been made of the factors that influence the accuracy of the results, especially the depolarizing effect of dust particles and of water vapor (thin clouds) in the earth's atmosphere.

Similar visual measurements have been made of the amounts of plane polarization in sun's rays after reflection by terrestrial materials of different kinds. The amount of plane polarization thus introduced depends not only on the phase angle or angle between the incident beam and the line of sight to the observer, but on the refractive and absorption indices of the material and on the nature of the surface that scatters the light, whether it is a surface of fracture, a rough or finely ground surface, or a coarse or fine powder surface. The factors that enter into the problem are many; their influence can be determined only by adequate numbers of measurements on each material under test.

The phenomena observed are chiefly those of scattered (diffracted) light combined with absorption and internal and external effects of reflection; under these conditions the character of the surfaces that scatter the light plays an extremely important part and gives rise to phenomena that cannot be explained on the basis of reflection and refraction alone. When the angle between the incident and the radiated beam is small, at some value between 10° and 30° depending on the substance and its surface, the amount of polarized light in the scattered beam is zero. For still smaller phase angles the scattered beam is again partially polarized, but the plane of vibration of the polarized component is in the plane of incidence rather than normal thereto.

Lyot, who was the first to observe this phenomenon in moonlight, called it negative polarization. It appears in moonlight about two days before full moon and continues for an equal period after full moon. Terrestrial materials exhibit the same phenomenon and approximately the same range of variation. The percentage amount of negative polarization rarely exceeds one per cent and may be less. In terrestrial materials it varies with the substance and with the character of the scattering surface. As a diagnostic feature negative polarization is of little value. It is, however, an important element in the theory of scattering of light and is being investigated further.

The visual measurements have been made with a special eyepiece consisting of a tilting-plate compensator together with a detector for ascertaining the point of exact compensation. The field of the eyepiece is a divided photometric field in which two factors, equality of illumination and exact alignment of Savart fringes, are the two criteria used in making a measurement; it is the combination of these two factors that renders the method so accurate. With this eyepiece the percentage amount of polarization can be determined with an accuracy of 0.2 per cent. The plate of the compensator can be tilted up to 70°; for this angle of tilt the percentage amount of polarization introduced by a single thin plate of glass or celluloid of refractive index 1.505 is approximately 26 per cent; at a 60° angle of tilt it is about 18 per cent; with a compensator consisting of two plates mounted in parallel and tilted at 60°, the percentage amount is increased to 30 per cent; with 4 plates, to 46 per cent; with 6 plates, to 56 per cent; and with 12 plates, to 71 per cent.

For observations on the moon the single-plate compensator suffices; the maximal polarization in moonlight from the dark areas or Mare is approximately 17 per cent, and from the bright lunar mountain areas it rarely exceeds 8 per cent. For observations on terrestrial materials more than one tilting plate is needed to compensate the much higher amounts, exceeding 60 per cent, of plane polarization produced by dark-colored rocks, such as basalts and peridotites.

With the aid of the polarization eyepiece, equipped with tilting compensators of one, two, and six plates, polarization measurements have been made on fifty selected terrestrial substances illuminated by sunlight incident at different angles and viewed from different directions.

The results of visual measurements of polarization in light scattered by the moon's surface and by terrestrial materials have been plotted on graphs and are now being assembled, preparatory to a report on this part of the determinative work.

Thermoelement. Many measurements have been made of the amount of plane polarization in a beam of light with the aid of the vacuum thermoelement together with a special rotatable Wollaston prism of quartz and a lens system of fused silica. The sensitivity of the thermoelement is not sufficiently high to be satisfactory for this purpose, but it extends over the entire spectrum and is useful as a check on measurements by other methods.

Polarization spectrograph. A spectrograph consisting of optical parts of fused silica and a Wollaston prism of quartz in a sliding mount has been used for polarization measurements and many spectrograms have been taken. In the parallel beam between the collimator and the first prism of the spectrograph the Wollaston prism can be inserted and two spectra obtained, the one with vibrations in the plane of incidence and the second with vibrations perpendicular thereto. With this arrangement each spectrogram has imprinted upon it three spectra, one above the other; of these the two outer spectra are the polarization spectra, and the central one is the spectrum of the entire beam. The polarization spectra are imprinted by a second exposure after insertion of the Wollaston prism. Difficulty was experienced in this arrangement with the calibration of the negative film in translating densities of the negative into intensities of incident light. To avoid this situation a tilting plate of thin mounted celluloid was built into the instrument. With it any definite percentage amount of polarization light can be introduced into the beam of incident light and the resulting changes in intensity measured on the spectra thus obtained, thus converting the method into a null method. The sensitivity of this method is not high but the spectrograms yield information both on the percentage polarization for any given wave length and on the relative intensities for different wave lengths. The polarization spectrograph covers a wide range of wave lengths, through the ultraviolet into the near infrared of the spectrum. Measurements with this method are being continued.

Photoelectric cell. Two different methods with the photoelectric cell as receiver have been employed. The first is based on the use of direct current and the second on alternating current. In the first attachment a large Wollaston prism of quartz in a rotatable mount is placed in front of a vacuum potassium Kunz photocell of fused quartz, the current from which is passed to the special amplifying circuit of DuBridge and Brown as adapted and improved by Dr. J. Stebbins. This method has been tested on a number of rock samples but has not yet been applied in series of routine measurements. The second method utilizes a mounted polarization prism that is rotated about its axis five times a second. If the incident beam of light contains a certain amount of polarized light, the intensity of the transmitted light will be a maximum whenever its plane of vibration coincides with that of the rotating polarizing prism, thus giving rise to an alternating current of 10 cycles per second in the photocell. A special high-gain amplifier for use with this arrangement has been devised by Mr. Ellis Johnson, of the Department of Terrestrial Magnetism; through the courtesy of Dr. J. A. Fleming the amplifier was built in the shop of that Department. The attachment has been thoroughly tested recently, but has not yet been employed in routine measurements. It gives promise of high sensitivity, but will require some modification before it is entirely satisfactory.

The foregoing methods have now been tested and their suitability for the task ascertained. Series of routine measurements with their aid are still outstanding.

The results of the visual measurements of the amount of plane polarization in sunlight reflected by lunar and terrestrial materials indicate that the substances exposed at the surface of the moon are of the nature of volcanic ashes and pumice high in silica. The changes in the amount of polarization in the reflected light for different angles between incident and reflected rays prove that the lunar surface is rough and not smooth. The changes in the intensity of light from different areas on the moon point to the same conclusion. It is true that the observations are restricted to materials at the actual surface of the moon and that in certain regions the cover of ashes and pumice may be quite thin and, like a light snowfall, may serve to blanket and conceal whatever is underneath. For this reason efforts have been made to measure the amount of plane polarization in sunlight reflected from steep lunar mountain slopes and inner crater walls that may not be coated with an insulating layer. These measurements are being continued with more powerful equipment.

Further evidence on the nature of lunar surface materials was obtained by Drs. Pettit and Nicholson during the lunar eclipses of June 14, 1927 and of November 26, 1928. Their measurements were made with the aid of a vacuum thermoelement and of screens of fluorite, glass, and a water cell mounted on a large reflecting telescope. The results showed that during an eclipse, when the moon enters the earth's shadow and is cut off from the sun's radiation, its surface chills so rapidly that, in the course of an hour or two, its temperature drops from +120° C. to approximately −100° C. From the data of measurement thus obtained Dr. P. Epstein has shown that the actual surface layer, which absorbs and radiates heat, is very thin; in other words, that lunar surface substances are good heat insulators, resembling, in thermal behavior, volcanic ash and pumice, or other highly porous substances. Large uncovered areas of massive rocks, such as granite, when heated to the boiling temperature of water and then set in a cool place require a long time to cool; they are thus precluded. If massive rocks occur near the moon's surface they are now covered by pumiceous material and well protected from the disintegrating effect of the alternations of extreme heat and cold to which the lunar surface is exposed.

Photographic Map of the Moon

Study of the surface features of the moon has emphasized the need for a lunar map that is free from the personal factors that have entered heretofore into the construction of maps of the moon. Of maps there are two kinds, the plan or base map and the topographic. Thus far only the first kind has been attempted for the moon. It represents the moon's surface projected on a definite plane and shows the features somewhat as the astronomer sees them through his telescope. These maps have been drawn by astronomers untrained in the principles of map making, with the the result that existing maps are unsatisfactory in several respects: the balance between map scale and amount of detail shown is not realized and some of the maps are not easily legible; several lunar maps have been prepared by men who were good observers but not good draughtsmen and unable adequately to portray what they saw. In other words the existing maps suffer from the personal equations of the men who prepared them. Comparison of a lunar map made a century ago with one drawn recently shows marked differences in the appearance of certain features; on the basis of such a comparison it has been concluded that changes have taken place here and there on the moon. But astronomers do not agree on the validity of any single change and the bulk of the available evidence goes to show that there has been no appreciable change on the moon's surface within the past century.

It seemed, therefore, to the Moon Committee that a lunar map should be prepared which is free from the personal equation and not dependent on the skill of the observer to depict correctly what he sees on the surface of the moon. The positions of approximately 4000 points on the moon's surface have been accurately measured by Saunders, Franz, and others and expressed in terms of selenographic longitude and latitude. With the aid of these data on position it is possible to ascertain the amount and direction of libration in each photograph of the moon. If each photograph could be transformed so that its plane would coincide with the plane of mean libration, namely, the plane on which all lunar maps are projected, the transformed photograph would form part of a lunar map and at the same time be free from the personal equation of the one who prepares the map.

To prepare a photographic map of the moon it is necessary to transform photographs taken with the aid of the 100-inch telescope at Mount Wilson so that the plane of projection is the same for all photographs. A map is a projection on a definite plane; the type of projection and the plane of projection must be quite definite if the map is to be satisfactory. For the transformation of the photographs a special moon house has been built at Mount Wilson. It is an insulated structure with double walls, corrugated sheet iron on the outside and paper on the inside with ventilation between the walls so that they quickly respond to temperature changes outside. The floor is covered with a layer of sawdust 6 inches thick to prevent radiation from the ground. As a result, the temperature distribution within the 150-foot building is remarkably uniform and seeing conditions are good so long as the temperature outside is not changing rapidly and there is no appreciable wind.

To transform a given moon photograph taken at the Cassegrain focus of the 100-inch telescope (focal length 135 feet), the moon positive, 15 inches in diameter, is mounted in front of a powerful beam of light reflected by an Army searchlight mirror 3 feet in diameter; the light passes through the positive to a parabolic silvered mirror of 67.5 feet focal length and 135 feet distant and thence back to a carefully turned globe of bronze, 15 inches in diameter and coated with magnesia powder. This coating furnishes a white diffuse reflecting surface. The image of the moon formed on it is in all respects similar to the moon in the relations of the surface features one to another; in other words, it is a miniature moon which can be photographed from any direction. For this purpose a second reflecting mirror, also of 67.5 feet focal length, is placed at such a position that it views the moon from the direction of mean libration and casts an image of it on a photographic plate mounted in a compartment beside the illuminated globe. The photographs thus produced are actually projections on the plane of mean libration; they fulfill the requirements of a map on a given scale. In order to complete the series of maps showing the moon at different phases, photographs of the moon, taken in July 1938 with the 100-inch telescope at the Newtonian focus, are to be transformed so that the plane of projection is that of mean libration for each negative. The series of maps made by this method will be free from the personal factor and of more value a century hence than at present.

Topographic map. A simple method was developed several years ago for ascertaining, from photographs of the moon, the slopes, shapes, and dimensions of its surface features. The moon is so nearly spherical in shape that, for all local measurements, it may be considered to be a sphere. A photograph of the moon is a projection of its spherical surface on a diametral plane normal to the line of sight. With the aid of an accurate lunar perspective polar projection of the same diameter as that of the lunar image on the photographic plate, it is possible, by superposing the projection on the photograph, to ascertain the angle which the incident sun's rays include with the normal to the moon's surface at any given point. If, for example, the inner slope of a crater is greater than the angle of elevation of the illuminating sun's rays, it will cast a shadow; if the slope angle is less, the surface will be illuminated; when slope angle and angle of elevation are equal, grazing incidence is observed.

From a series of photographs taken at different times during a lunation it is possible to observe the changes in character of illumination of any given slope and to ascertain its angle. In 1929 a series of motion pictures was taken of the moon with the 100-inch telescope on films in a motion picture camera. In July 1938 more than 400 photographs of the moon were taken at the Newtonian focus of the 100-inch telescope. A new 3-element zero corrector lens, designed by Dr. F. E. Ross, was used and found to be excellent for the spectral range between 5000 and 6000 Ångström units. This set of photographs will afford the basis for a better physiographic study of the moon than has heretofore been possible and will make possible the preparation of a rough topographic map of the central part of the moon's disk. The dimensions of a crater, for example, are read off directly on the superposed projection chart in terms of latitude and longitude angles or of polar distance and azimuth angles; these are, in turn, reduced to linear distances by use of a conversion table. With the aid of the Nautical Almanac it is possible to ascertain the phase angle (angle at the moon's center between the lines of sight to the centers of the sun and earth respectively) for the instant of time at which a given photograph was taken. With the aid of the projection the angle of elevation of the sun's rays at any given point on the moon's surface and for any given phase angle can be read off directly without computation. By this method the observer is in position with a sufficiently complete set of photographs to reconstruct the shapes, both in plan and section, of lunar surface feature within 45° of the center of the moon's disk and to draw therefrom a rough topographic map. Work on the preparation of this map will be started as soon as prints from the negatives have been made. Preliminary measurements of slope angles on negatives taken at the Cassegrain focus of the 100-inch telescope prove that the inner slopes of small craters (craterlets) are greater than those of the larger craters and walled plains. The slope angle of the inner wall of some of the smaller craters ranges from 50° to 55°.

Photograph of Moon on Globe

The projection of a moon positive on the magnesia-coated bronze globe in the moon house gives an extremely beautiful and realistic presentation of the moon's surface. The undistorted appearance of the craters and other features near the edge of the moon's disk is helpful in the visualization of the surface relationships. In order to make this globular relationship more accessible, a glass globe coated on the outside with photographic emulsion was substituted for the magnesia-coated bronze globe and the moon negative projected on it, thus producing a moon transparency which is angle true. The globe is frosted on the inside and illuminated by an electric lamp. The coating with photographic emulsion was done, through the courtesy of Dr. C. E. K. Mees, by the Research Laboratory of the Eastman Kodak Company, and represents an advance in photographic technique. A number of these globes have been prepared and will be useful to the Moon Committee in its physiographic studies later; they serve also as exhibits of miniature moons, 13.5 inches in diameter, and show the aspect of the moon at different phases.

Lunar Surface Features

Although the Moon Committee has thus far devoted the major part of its available time to polarization measurements in the effort to ascertain the nature of the materials exposed at the moon's surface, it has made many observations on the surface features themselves and has sought to interpret some of the extraordinary features in terms of the conditions known to exist at the lunar surface, such as absence of atmosphere, of water and water vapor, and of the effects of ordinary erosion by running water; gravity one-sixth of that of the earth; rapid changes in temperature from +120° C. at noon to less than −100° C. at midnight. It has made a detailed study of the trajectories of materials hurled out of lunar craters and found that they are from 20 to 50 times longer than those of materials ejected at the same initial velocity and angle of elevation on the earth.

Under such conditions materials ejected from a lunar crater are scattered far and wide, whereas on the earth the greater part of the ejected rock fragments fall near and into the crater orifice. Lunar craters are cleaned out, as a rule, and are of the nature of immense deep holes in the ground with the floor of the crater below the level of the surrounding country; the floors of terrestrial craters, on the other hand, are near the top of the crater and high above the level of the adjacent country. On the moon, few, if any, typical lava flows have been observed. The floors of the latest craters are commonly whiter and brighter than older craters; this is especially true of the craterlets and of the "rays" from many craters. The structural features follow in many areas definite patterns; along many structural lines rows of craterlets are observed, as also craterlets in the central hills of larger craters. These relationships occur in too many regions on the moon to be ascribed to mere chance and indicate strongly that many of the craterlets are of volcanic rather than meteorite impact origin. On the other hand, the translational energy of a meteorite impinging unimpeded on the moon with a velocity of 20 to 40 kilometers a second and penetrating into its surface for some distance is able not only to produce the crater form but, on transformation of its residual kinetic energy into heat, to melt and even to volatilize the invaded rock and thus to give rise to actions that in their effects closely resemble volcanic phenomena.

Study of the "rays" radiating from craters, such as Tycho, has shown that they do not project as ridges above the land they traverse; they cast no appreciable shadows when near the terminator; they reflect the light more strongly than do the materials of the surfaces over which they reach. It was suggested many years ago by different observers that the rays from a crater are streaks of fine dust or ashes blown out from its top or sides, and carried along by the jets of hot escaping gases from which they gradually settled out along the paths followed by the streaming gases. This explanation of the origin of the rays seems to accord well with the observed facts. The rays, like many of the craterlets, are younger than the land on which they occur. Observations by the Committee indicate clearly that the floors of younger craters are whiter and brighter than those of older craters and that, as a general rule, the degree of whiteness of a crater floor is an indication of its relative age; it appears that the younger craters have brought to the surface white fresh materials that have not yet been darkened by the staining action of slowly acting weathering agencies and by the gradual accumulation of a thin mantle of cosmic dust.

A preliminary study has been made of the relative frequency of lunar craters of different diameters, as listed in the publication entitled *Named Lunar Formations* by M. A. Blagg and K. Mueller of the International Astronomical Union, 1935. The total number of craters for which the diameters are given is approximately 3950. The analysis shows that the distribution curve is remarkably smooth and approximately exponential in character. No satisfactory explanation has been found for the type of the frequency curve.

Program for Future Work

The work of the Committee on Study of the Surface Features of the Moon has progressed relatively slowly. Each member of the committee has had but little time to give to it; the Committee realized early in the task that, before physiographic studies of value could be attempted, more definite information was needed regarding the nature and environment of the lunar surface materials; also that a topographic map or a simple method for ascertaining the shapes and spatial relationships of the various surface elements was desirable. These tasks had of necessity to be undertaken not by the whole Committee but by one or two of its members, working at Mount Wilson Observatory for a month or so each year and calling upon the other members for advice whenever needed. This deliberate approach to the problem has had the advantage that it has enabled the Committee to plan the attack well and to carry it out with relatively small cost, even though the work is spread over a number of years.

The Committee has developed new methods that are yielding results of measurement of known accuracy, thus providing a firm foundation for the critical analysis of the data obtained. At the present rate of progress several years will be required to finish the measurements on polarization by the different independent methods and to complete the topographic map and the series of photographic maps. It will then be possible for the committee to proceed to the problem of physiographic description and interpretation of the surface elements of the moon with the assurance that the available data are reliable and can be trusted within definitely determined limits.

The policy of assigning to an interdepartmental committee a problem of large scope is in keeping with the general policy of the Carnegie Institution of supporting organized efforts in fields of science too large for one man to encompass. Each department of the Institution is essentially a group of cooperating scientists, each member of which is engaged partly in researches of his own and partly in the attack on larger problems in cooperation with other members. This group method of facing each problem from all standpoints and of determining the best means for solving it has, in recent years, been extended by Dr. Merriam to cooperative work between the several groups within the Institution and between the Institution and outside groups.

The dividends accruing from cooperative work of this type, in terms of scientific results obtained for a given sum of money, are large, chiefly because of the facilities and background of experience within the several groups. Emphasis by Dr. Merriam on the unity of the Institution in its approach to new problems in science and in its relations with the public has been an extremely important factor in increasing the effectiveness of its work, and in stimulating the interest of each staff member in fields of research other than his own. This shift in attitude has led to better appreciation of the meaning and value of cooperation in research in science.

THE ELECTRIC EYE IN ASTRONOMY

JOEL STEBBINS

Director Washburn Observatory, University of Wisconsin
Research Associate, The Carnegie Institution of Washington

THERE is no more striking illustration of the dependence of applied science upon pure science than the photoelectric cell or so-called electric eye. In one form or other this device consists of a metallic surface in an evacuated glass tube, which on exposure to light releases electrons constituting an electric current which may by suitable magnification be applied to purposes as widely varied as human imagination will suggest. The photocell is the basis of the modern talking picture, of television, and of the transmission of photographs by radio or by wire. It is used for sorting beans, buttons, and cigars, for counting automobiles and pedestrians, for timing horse races, for stopping or starting anything, even up to a World's Fair, and finally for measuring the size of the Milky Way.

It was the advent of the thermionic amplifying tube, developed for the radio, that made possible the amplification of the minute electrical current in a photocell up to a size that could be applied to operating relays or to controlling still larger currents. About ten years ago the silent motion pictures were made over into the talkies by means of the photocell, but the same sort of cell has been used for measuring the light of stars for a quarter of a century. In the early nineties Dr. George E. Hale used a photocell on the summit of Mount Etna in an attempt to detect the solar corona without an eclipse. He used a cell which had been developed by Elster and Geitel, and they in turn were preceded by Hallwachs, who had found that negative electricity would leak from a conductor when it is exposed to ultraviolet light. Thus a phenomenon discovered in the laboratory was applied by the physicist and astronomer to the problems of pure science for nearly fifty years before it was taken up for the widespread commercial uses of today.

In measuring or detecting faint lights with a photocell there is in principle no better method than to connect the sensitive metal of the cell with a wire to the needle of an electrometer. Then as the connected conductors take on a positive charge owing to the loss of negative electrons from the surface of the metal, the swinging end of the needle is drawn toward a charged plate just as a pithball is drawn toward a rod in the first simple high school experiment in electricity. The manipulation of a sensitive electrometer requires great experimental skill, and it is easier to apply the special thermionic tubes, or so-called electrometer tubes, developed by the General Electric and the Western Electric companies. Even these tubes work most satisfactorily when both the photocell and the amplifier are enclosed in an evacuated tank with a window for the entrance of light. In this arrangement the effect of leakage caused by moisture on the insulators or of stray charges produced by cosmic rays passing through the air are eliminated, and it is literally true that the effective sensitivity of a photocell can be increased ten- or a hundredfold by enclosing it and the electrometer tube in an air-tight metal container.

Such a photometer perfected by Dr. A. E. Whitford and used at the Washburn and Mount Wilson observatories is shown in plate 1, figure 1. The cell and tube are mounted on a brass base fitting the tank, which in turn is readily attached to the eye end of a telescope. The sensitivity of this arrangement is shown by an experiment which we performed at Madison. For many years we had known by calculation that the photocell could detect a candle at a distance of several miles, so finally on a dark night we set up a real candle at the actual distance of a mile from the Observatory, on Picnic Point across Lake Mendota. The cell without any telescope was pointed in the right direction, and when the candle was lighted the resulting amplified current gave a deflection of the galvanometer of more than 150 divisions when a single division could have been detected. A weaker light, about one-fiftieth of the candle at one mile, was readily measured, and it was found that the practical limit of detection was not far from a thousandth of a candle at one mile, or a whole candle at thirty miles. Since the limit for the human eye is a candle at five or six miles, the photocell thus shows up as an extraordinarily sensitive receiver of radiation. The cell had an effective diameter of one inch, and therefore had an area ten or fifteen times that of the pupil of an eye in the dark. When reduced to the same sized opening the relative sensitivities of the eye and photocell would depend upon the color of the light source; some cells would be more sensitive than the eye for a blue light, while the eye would be better for a yellow one.

The foregoing experiment is useful in indicating the sensitivity required in applying the photocell to measures of faint stars. A candle a mile away looks about as bright as the Pole Star does to the unaided eye. Hence to find out whether a given set-up in the laboratory will work on the telescope one need only point a cell out of a north window at Polaris. This is exactly what we did in the early experiments at the Washburn Observatory. The tank containing a photocell and amplifier was placed on an inclined board; then the cell was directed through a blank tube at the star. Exposures were made with a rough shutter on the sky and on the star, and from the resulting galvanometer deflections we knew what the sensitivity would be for faint stars. The object glass of the $15\frac{1}{2}$-inch refractor collects 240 times as much light as a 1-inch cell, and hence would enable us to reach stars six magnitudes fainter than with the cell alone. For the past eight years we have used a photoelectric photometer on the large reflectors at Mount Wilson, and the faintest objects measured with the 100-inch are of the sixteenth magnitude or forty times fainter at the cell than Polaris is without a telescope.

We also have made direct comparisons of the relative sensitivity of the photocell and the photographic plate. Photographs of a star were taken with different exposures, and then with the same lens the star was measured with a photocell and a fast recording galvanometer. The photographic images were discernible for 1/50 second but not for 1/100 second. The cell, however, always gave a measurable record for an exposure of 1/100 second. Therefore, for a single white star of the color of Sirius the photocell in a given time is at least twice as sensitive as a photographic plate. However, the cell is all through in two or three minutes, while the plate keeps on building up an image for hours. Also, while the cell is receiving the light from only one star the plate can record hundreds or thousands at the same time, and there is the very great advantage of the permanent record of the photograph. Likewise in the case of stellar spectra, which astronomers spend many weary hours in photographing, a photocell will record only one small region of a spectrum while the photograph is recording in separate detail all the colors from red to blue or whatever regions the plate will take in. Thus there is no prospect that the photocell will supplant the photographic plate in investigations where more than one star image or more than one spectrum line is to be studied.

The photocell does however have an advantage in the measurement of surface brightness of extended areas, where all parts of the image at the focus of a telescope are added up into one total magnitude. It may require an hour to photograph the faint night sky which can be recorded and measured in half a minute with a photocell. Since some of the extended nebulosities are difficult to disentangle from the sky background, the photocell can well be applied to the study of these objects. It was in just this measurement of the luminosity, or, putting it the other way, the blackness of the night sky, that the photocell could make a contribution in the selection of the site for the 200-inch telescope on Palomar Mountain. One reason for not placing this telescope on Mount Wilson was that the increasing illumination of the sky by the lights in the Los Angeles area has been interfering with the long-exposure photographs for a number of years. The amount of this illumination has been measured with a photocell. With a portable 3-inch telescope the brightness of the sky near the pole was measured, using Polaris as the standard. The instrument was set up for several nights first at Mount Wilson, then at Palomar, and again at Mount Wilson; such series being taken in both summer and winter. On the best nights at Mount Wilson, when fog was over Los Angeles and other cities, the sky was nearly as dark as it was at Palomar. Usually, however, the sky at Mount Wilson, even toward the north and away from the city lights, showed a detectable illumination, which was greatest in the early evening and which gradually decreased as many of the lights of the communities were extinguished toward midnight. The darkest sky at Mount Wilson was about 25 per cent brighter than the darkest sky at Palomar, which means that photographic exposures could be continued 25 per cent longer at Palomar before the illumination would affect the plates to the same extent. So far as can be determined, the sky at Palomar is as dark as can be found in any practicable site in this country. A wholly dark sky is not to be found on the earth, as there is always the faint glow of the permanent aurora which is tens or scores of miles above the earth's surface. Also the faint illumination of the air by all the stars and the general zodiacal light are always in evidence.

Another application of the photocell to a quantitive measurement of the steadiness of the air, or the "seeing," was developed too late to be of service in the selection of the site for the 200-inch. In the search for the best possible location a number of observers with small telescopes at different sites made continuous visual observations throughout many nights, and their results were reduced to a common system devised by Dr. J. A. Anderson. During the past year Whitford has perfected a combination of photocell and alternating current amplifier which with a 6-inch telescope gives a scale reading for the variation of brightness or twinkling of a bright star. The changing brightness can also be shown and recorded with a cathode-ray oscillograph. In plate 2, figure 1, are shown the variations in the light of a star produced by disturbances or waves in the atmosphere.

It is true that here we have a measure of only part of the imperfections of a star image at the focus of a telescope. Astronomers recognize two kinds of bad seeing, one where the image is small or "hard" but wanders around, and the other where the image is blurred or "soft" but is too large to show much motion. The atmospheric waves or disturbances for the first kind of seeing may be considered to have dimensions to be measured in feet, while for the second the size is a matter of inches. Accompanying the jumping or blurring of the image is the variation of intensity, or the twinkling which is visible to the naked eye. It is this twinkling which is detected by the photocell, and not the wandering or blurring of the image. In plate 2, figure 1, the records of a star in good and bad seeing show the great changes in intensity within a fraction of a second, and it was something of a surprise to us to find how much the light of a star image in a telescope actually does vary. We have many measures of stellar brightness correct to less than 1 per cent, but here the star was changing by 50 per cent or more. Our ordinary photometric measures are based on galvanometer readings requiring ten or fifteen seconds for a scale deflection, which smoothes out the short-period jumps, and with a large telescope the effects of the air waves are also reduced. Nevertheless, even with the 60-inch or the 100-inch telescope a wandering of the galvanometer deflection by 1 per cent or more can be observed when a photocell is exposed steadily to a star. The conclusion is that precise photometric measures of stars must not be taken too rapidly because of the variable effect of seeing.

An extension of the technique of measuring the twinkling of stars was worked out by Whitford and G. E. Kron in the development of an automatic guider for photographic telescopes. With all large telescopes the photographic exposures are made by the astronomer keeping a star image accurately centered on the intersection of two cross wires, sometimes for hours. The main driving clock follows closely the diurnal motion of the stars, but no mechanical device can make up for the wandering of a star image due to bad seeing. Therefore the observer has to attend continuously to the exact pointing, and by turning a screw, sometimes several times per second, he moves either the telescope as a whole or a sliding plateholder to keep the star properly centered. This laborious guiding is the bugbear of astronomical photography, and in the long night hours every observer has had the opportunity to invent a machine to do the work. The standard idea, which probably has occurred to almost everyone, is to let the star image fall on the apex of a four-sided reflecting glass pyramid. When the image gets off center and slides down one face of the pyramid, the light will be reflected to a photocell; then with proper relays a small motor or magnet will move the pyramid and plateholder until the star image is central on the apex once more. Barring the fact that four photocells are required, the idea is simplicity itself; all that is necessary is to make it work. If the light from a star were strong enough, there would be little difficulty in carrying out the scheme, but every inventor promptly finds out that the stars are pretty weak sources of power. The ordinary automatic device which works with a photocell usually requires an illumination at the cell equivalent to the light from a candle at a distance of two or three feet. Now move the candle a thousand times farther away, where it becomes a million times fainter, and you have the condition for an automatic guider for a telescope. Actually the stars at the focus of a telescope lens are nearer ten million than one million times fainter than a candle at a few feet.

In their first experiments with the automatic guider Whitford and Kron worked in one coordinate only, the east and west motion, which is the more important, and they substituted a roof prism for the pyramid. With the edge of the prism running north and south, the light beam from a star was divided equally and reflected by proper mirrors to a single photocell. Then with a rotating sector or "light-chopper" the two beams were alternately occulted, and with the star in position there would be no flicker of intensity on the photocell. With the star over on one face of the prism one beam would be stronger and there would be a fluctuation or flicker at ten cycles per second, the period of the light-chopper. The resulting effect was magnified by an alternating current amplifier controlling a small motor which would drive the sliding plateholder in the direction to correct the error. The complete outfit contained half a dozen amplifying tubes, and was not much more complicated than the ordinary radio receiving set.

After some months of experimentation at Madison the guider was taken to Mount Wilson, where the first actual photographs were made. Figure 2, plate 1, shows a photograph taken with the 10-inch telescope on the first night that the automatic guider was tried. The telescope was set to run fast so as to give the guider something to do, and the resulting star images were as good as could be gotten by manual guiding. It is true that the exposures were only ten minutes long, but the first ten minutes are the hardest. The guider was later attached to the 60-inch telescope and satisfactory photographs were taken with guiding stars as faint as magnitude 8.5. It is easy to compute that a star of that magnitude at the focus of the 60-inch is equivalent to a candle at the distance of half a mile from a cell with a 1-inch opening, but even stars of this brightness are not so numerous in the sky that one would usually be available in the small area in the focal plane of the telescope at one side of the plateholder. To be generally useful with the 60-inch the guider ought to work on stars of magnitude 10, or four times fainter than those of magnitude 8.5. In other words, the guider should work with a 1-inch telescope on a candle a full mile rather than half a mile away.

In Whitford and Kron's device the electron multiplier of Dr. V. K. Zworykin was used instead of an ordinary photocell. This multiplier gives a millionfold amplification of the photoelectric current within the original tube with very little additional "noise." The power amplification from the first emission of electrons up to the current which drives the motor of the guider is of the order of 10^{18}, so it is evident that some care must be taken not to introduce disturbances somewhere along the line. It is not clear just how valuable an auxiliary the guider will turn out to be for stars, but for work on the sun Whitford has already constructed a guider which follows better than can be done by hand.

Just after the foregoing paragraphs were written, the *Observatory* for May 1938 was received, containing the following quotation from fifty years ago.

> MEETING OF THE ROYAL ASTRONOMICAL SOCIETY, 1888, APRIL 13
>
> *Mr. Common.* It struck me when listening to Sir Howard Grubb's remarks in explanation of the working of this clock, that if we had some means of using the image of the star itself to regulate the motion, we should get a perfect clock. If we could find a substance that would vary in its conductivity sufficiently through the impact of light, as selenium does to some slight amount, and we placed the image of the star between two bars in such a way that if it touched one bar it would retard the clock and if it touched the other bar it would accelerate it, we would then have an ideally perfect clock.

Another use for the photocell in the taking of astronomical photographs could be in the form of an exposure timer. We are all familiar with the exposure meters which now can be had in any camera store and which take most of the uncertainty out of amateur photography. In photographing stellar spectra the amount of light from a star which gets through the slit of a spectrograph is very much dependent upon the seeing, and the observer has to make a good guess on the exposure time. It should not be difficult to place a photocell in position to catch some wasted light which is reflected from the face of a prism or grating, and which could be measured as the exposure is going on. As the time with a large telescope is precious, it would even be worth while with the 200-inch reflector to assign some young physicist the job of telling the astronomers the proper exposure times for their spectrograms. It would probably not be long before the physicist would have a device to do his work for him automatically.

We now turn from these experimental or projected applications of the photocell to fields in astronomy where the cell has been of service for many years. The basis for nearly all estimates of our own and other galaxies, even of the size of the universe itself, is a photometric one. When the stars are too far away to be measured for parallax or motion we have to rely upon the apparent magnitude to get the clue to the distance. The present system of magnitudes, from the first to the twenty-first, covers a range of more than one hundred millionfold in apparent luminosity, and it is a problem to set up standards of brightness over such a range which will be correct to even 10 or 20 per cent for the faintest stars and nebulæ. A beginning has been made with the photocell for checking the photographic magnitudes in the vicinity of the North Pole, and it is probable that in the future the cell will be used more and more to establish primary standards to which the photographic magnitudes will be referred.

During the past twenty-five years the most useful application of the photocell to stellar photometry has been in the detection of variations in the light of individual stars, variations too small to be noticed either by the eye or by the photographic plate. As an illustration, at Madison we have found that practically all the very red stars are intrinsically variable to the extent of at least 10 per cent of their total brightness. It is well known that stars like Betelgeuse in Orion present a variable brightness to the eye, but it is just these red stars which are most difficult to estimate visually because their colors are different from those of surrounding stars used for comparison.

Then there are the Cepheid or pulsating stars which vary in regular periods, and which, from the known correlation between length of period and total intrinsic luminosity, serve as standard lights for the estimates of distance when all other methods fail. Studies of these stars, which vary in cycles from a few hours up to several months, have given us a great deal of indirect information about the scale of cosmic distances, even though the mechanism which produces the light variation of the stars is only imperfectly understood.

But perhaps the most striking result of photoelectric photometry has been the detection of close companion bodies to the bright stars, which are revealed by eclipses when the companions are in front of the main stars. Generally speaking, such eclipses are easiest to detect when the faint companions are as large as or even larger than the bright bodies, and it has been the province of the photocell to detect the minute variation of light in cases where small companions cause the eclipses. Such a case is illustrated in plate 2, figure 2. The star AR Cassiopeiæ was known as a spectroscopic double star from the periodic shift of the spectral lines, and at the predicted time, when the companion was in front of the main star, the total light of the system was found to be diminished by about one-ninth, showing that the companion is $\frac{1}{9}$ the area or $\frac{1}{3}$ the diameter of the primary. When the companion is behind the main star, the loss of light is only $\frac{1}{4}$ of $\frac{1}{9}$; hence the surface brightness of the small star is $\frac{1}{4}$ that of the large one, and it gives less than $\frac{1}{30}$ of the total light of the system. Since the bright star is of the class that gives several hundred times as much light as the sun, the small star is only relatively a dark body, giving off probably ten times as much light as the sun. The companion of AR Cassiopeiæ is one of the very smallest attendants that has been discovered among the stars, and it is not a dark planet but a sun on its own account. As seen from another star properly placed, the planet Jupiter might cause an eclipse of the sun to the extent of only 1 per cent of the sun's light once in twelve years. So while we have knowledge of many large companions to other stars, there is at present no possibility of detecting small planets like those of the solar system.

Another application of the photocell is to the precise measurement of the colors of stars. Even to the naked eye there are differences of color among the bright stars. Sirius is called white, Capella yellow, Aldebaran and Betelgeuse red. Corresponding to these colors are definite characteristics of the spectral lines; in fact a star's spectrum gives a better measure of its color than can usually be made by visual or photographic means. If there were no loss of light by transmission through space, a star would appear to be the same color at all distances, and its apparent brightness would depend simply upon the law of the inverse square of the distance. It is found, however, that for stars at only moderate distances in our galaxy there is both a general and a selective absorption of light. It is natural to assume that such loss is caused by material particles which may be in the form of gas, dust, or chunks. Chunks or coarse dust would weaken the light of a star without changing its color, just as a thin cloud of water or ice particles only dims the light of the sun. Fine dust or gas molecules would cause what is called space reddening, just as the air molecules produce a reddening of the sun at sunset, an effect which is sometimes much enhanced by fine dust which gives us a really red sun. Large particles merely obstruct the light, small ones scatter the light, blue much more than red, so that when we look at a star through gas or fine dust much of the blue light is missing and the star appears yellow or red. Therefore if we know the intrinsic color of a star, the apparent color will give a measure of the scattering dark material between us and the star.

The bright blue-white stars of Orion or of the Pleiades are seen in their true color, and the characteristics of their spectra show that they are very hot stars, much hotter than the sun. Many similar and more distant stars, located along the central line of the Milky Way, are yellow or even reddish in color. These Orion stars, those of spectral class B, are recognized by their spectrum lines, and since they are of great intrinsic luminosity, from one to two thousand times as bright as the sun, they may be seen at great distances, and they make convenient objects for tests of the absorption of light in space. Such a test with the photocell is shown in plate 3, figure 1. There are many B stars in the star clouds of the Milky Way, but even where the general background of the stars looks brightest the B stars are not normal in color but show a distinct reddening.

It has been a matter of some technical difficulty to derive the proper numerical factor for changing the amount of reddening measured with the photocell into the corresponding total absorption of the light which affects the photographic plate, but the present result is that this factor is best represented by the figure 9. For instance, if a star is found to be 0.10 magnitude redder than normal on the photoelectric scale, the total photographic absorption is 9 × 0.10 or 0.90 magnitude. This absorption means the loss of about one-half of the light of the star, and since all the B stars in the bright star clouds of figure 1, plate 3, are colored at least by this amount, there must be a veil of absorbing material or dust between us and the brightest parts of the Milky Way. The reddest B stars thus far observed have colors indicating a total absorption of four or five magnitudes. Hence these stars shine with only about 2 per cent of their real brightness, and they are much nearer than would be inferred from their apparent magnitude.

The factor of reduction from measured color to total absorption is a figure of great astronomical importance. If the distribution of different-sized particles is such that the ratio of the selective to the total absorption is the same in all parts of space, then we have a simple means of correcting the inferred distances of stars by measuring their colors. The end is not yet; it is conceivable that the relative amount of scattering material may be widely variable in different parts of space, but wherever we find a reddening of a star there must be some general absorption and the star must appear fainter for its distance than it actually is.

Even better than the very luminous B stars for tests of space absorption are the globular clusters, composed of thousands or even scores of thousands of stars. The photocell will give the color of any light that is fed into it, and fortunately it is found that the globular clusters have a small dispersion in color, which is much like the color of the sun. The colors of some seventy of the approximately one hundred known globular clusters have been measured with a photocell on the 100-inch reflector at Mount Wilson, and it is found that the clusters nearest the line of the Milky Way are even more strongly reddened than the B stars. In plate 3, figure 2, the photographs by Baade on a blue- and a red-sensitive plate show that not only the cluster but most of the stars are affected by space reddening. The reddening of this cluster was discovered with the photocell, but the photographs have enabled Baade to demonstrate that the whole region of stars is affected by dark material which must be in the foreground.

The fact that both the reddest B stars and the reddest clusters are found near the median line of the Milky Way leads to the inference that there is a thin stratum of dark material near the main plane of our galaxy, as is found in the extragalactic nebulæ now supposed to be galaxies like our own. When such external systems are inclined so that we see them edge on as in plate 4, figure 1, it is quite common to find the dark streaks analogous to the break in our Milky Way. Formerly these dark bands were considered to be at the perimeters of rotating systems, but now that such dark material is recognized right in the midst of the stars about the sun it is agreed that the material extends into the interior of each system.

Our galaxy is considered to be a flattened system of some billions of stars concentric with the nearly spherical system of globular clusters. The dimensions of the combined system depend upon the distances of globular clusters which Shapley derived from the apparent magnitudes of variable and other stars in those clusters, and any absorption of light in space will affect such inferred distances. The largest amount of space reddening for a globular cluster is 0.8 magnitude on the photoelectric scale, which multiplied by the factor 9 gives seven magnitudes for the total photographic absorption. Therefore such a cluster is less than $\frac{1}{600}$ as bright as it would be if space were transparent. Without emphasizing these particular figures it is evident that corrections must be applied to all distances dependent upon the brightness of stars or clusters which are affected by absorption. These corrections are largest in the direction of the Milky Way, but become almost negligible at right angles to that direction.

The outcome of the allowance for absorption is to reduce the galaxy to something like one-half the size it was inferred to be before the absorption was recognized and allowed for. Instead of an over-all diameter of two hundred thousand light years it seems that one hundred thousand light years will take in most of the stars of the system. Figure 2, plate 4, gives a good representation of our present idea of the galaxy, but of course the true system is much more complicated. The stars and the dark material are more bunched or spotted, probably as in the system of figure 1.

The photocell has not only served in reducing the estimated size of our own galaxy, but it has also served in extending the apparent dimensions of other galaxies like the near ones in Andromeda and Triangulum. On photographs the boundaries of these nebulæ fade away into the sky background, but the photocell can detect small differences between sky and sky plus nebula which are lost on the plates. It is not surprising to find the extension in size of other galaxies, for it is not difficult to show that, seen from the outside, our own galaxy in the region of the sun would be very faint. The center of the galaxy might look like the nucleus of an ordinary spiral nebula, but the sun is in the outskirts of our system. Since we can look out at right angles to the galactic plane and see other galaxies, an outside observer could look in; probably he could look past us and on through to galaxies on the far side.

A large field of investigation remains in the study of brightness and colors of other galaxies. One definite result from photoelectric observations is that light is changed or obstructed very little if at all in transmission through the spaces between the galaxies. Whereas space reddening can be detected at distances of only a few thousand light years within our own galaxy, no such effect is found at distances measured by millions of light years in the spaces between the galaxies. Light passing through our system from the direction of the galactic pole is reddened and absorbed about as much as in passing through the earth's atmosphere, but according to present evidence there is no appreciable absorption in the light from the most distant nebula until it reaches the confines of our own galaxy. Naturally, there may be dark material within the nebula which prevents some of the light from getting out, but once free from this entanglement the light passes on unhindered to our own or another galaxy.

From the foregoing examples it is seen that the photoelectric cell has a wide application to astronomical problems. There are other devices for the electrical measurement of radiation, such as the selenium cell, the thermocouple, the bolometer, and the radiometer, but up to now the photocell in connection with the amplifying tube is the most sensitive for faint stars. Moreover the photocell is still in the stage of continual improvement. The human eye has presumably not increased in sensitivity in a thousand years, the photographic plate is changing slowly from decade to decade, but the photocell can be counted on for something new from year to year.

In the work in photoelectric photometry the present writer has had the collaboration of various colleagues and the support of different institutions. In particular, Dr. Jacob Kunz of the University of Illinois has produced photoelectric cells which, despite all the developments in scientific and industrial laboratories, still hold the record for sensitivity to faint lights. Dr. A. E. Whitford of the University of Wisconsin has developed and improved the use of the thermionic amplifier, and has continually collaborated in the application to new problems. Dr. C. M. Huffer has carried on a large share of the observations at the Washburn Observatory. We have had continued cooperation by the Departments of Physics of the University of Wisconsin and the University of Illinois. It was the action of the Carnegie Institution of Washington which made it possible to apply the photocell with the large telescopes at the Mount Wilson Observatory, and additional support has been given by the California Institute of Technology in connection with plans for the 200-inch reflector. Grants or assistance in the past decade have come from the National Academy of Sciences, the National Research Council, the American Philosophical Society, and the American Academy of Arts and Sciences. Also the great industrial laboratories have given the benefit of their latest developments, often without price. They include the General Electric Company, the Western Electric Company, the Bell Telephone Laboratories, the Radio Corporation of America, and the Westinghouse Electric Company. And finally there has been continual support by the Research Committee of the University of Wisconsin and the Wisconsin Alumni Research Foundation. Any success in the past and any improvement in the future will have been due to each and all of these agencies.

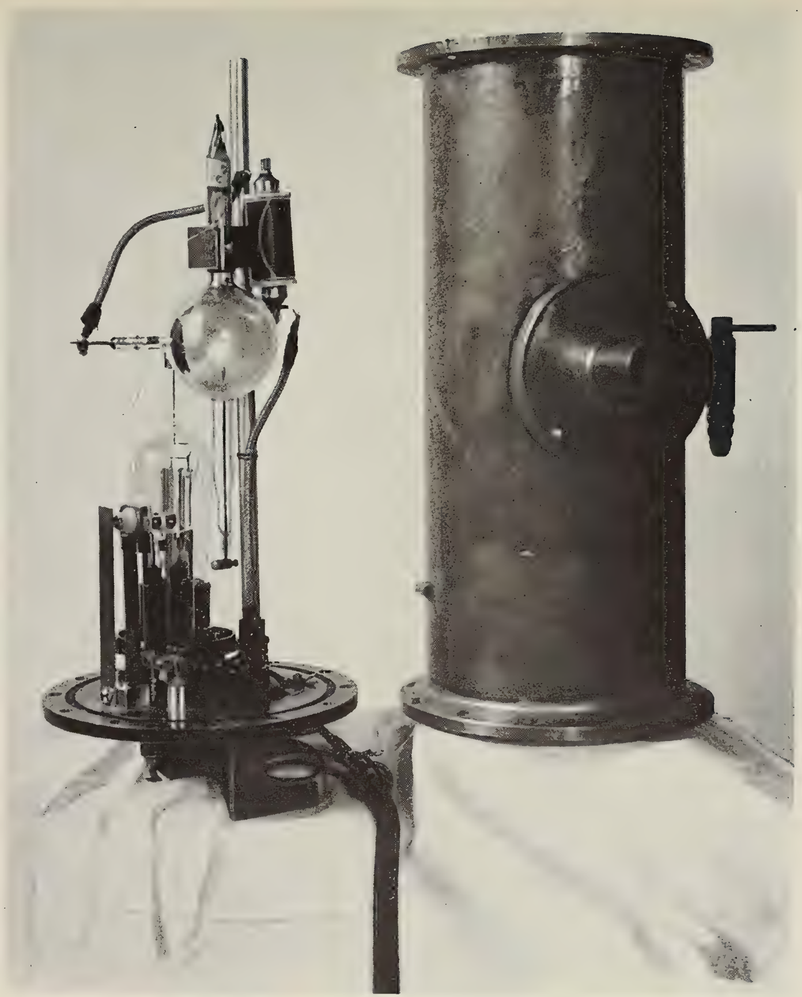


FIG. 1. Photoelectric cell and amplifier. Diameter of tank, 5 inches.

FIG. 2. Enlarged star images. With and without the automatic guider.

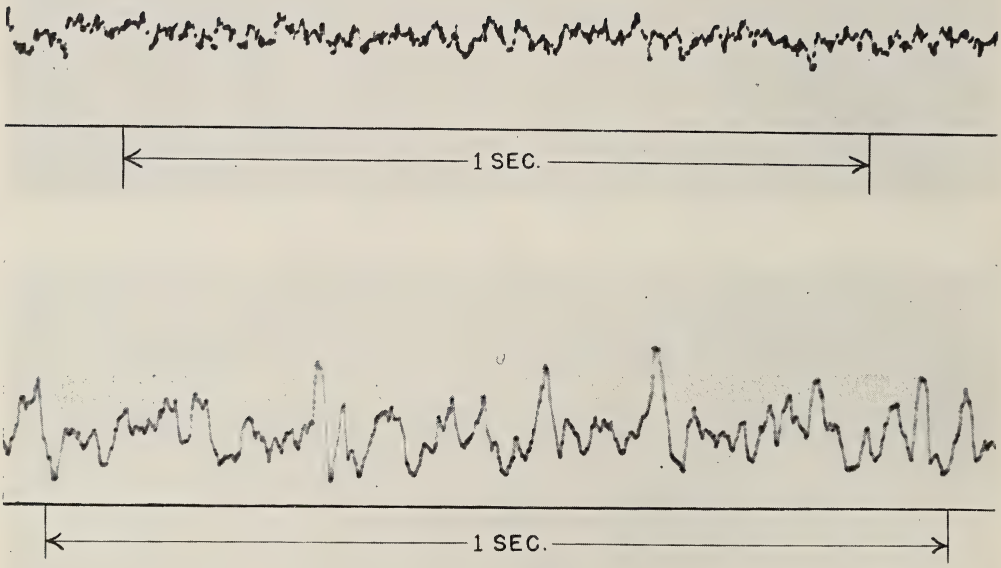


Fig. 1. Twinkling of star in good and bad seeing with 6-inch telescope. Intensities measured from horizontal lines.

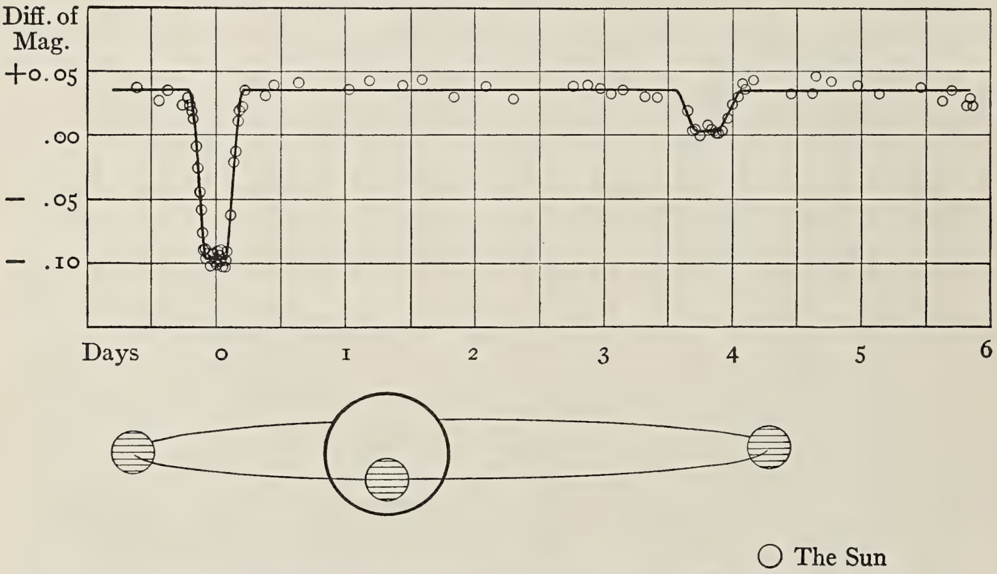


Fig. 2. Light curve and system of the eclipsing binary, AR Cassiopeiæ.

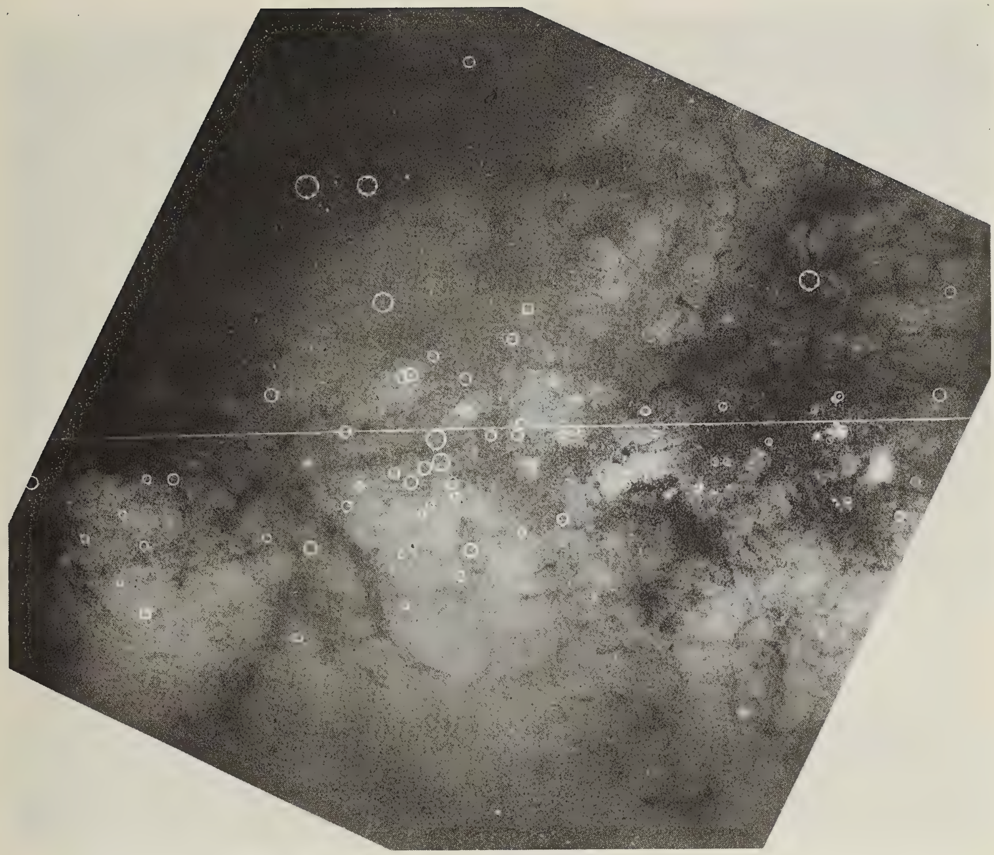
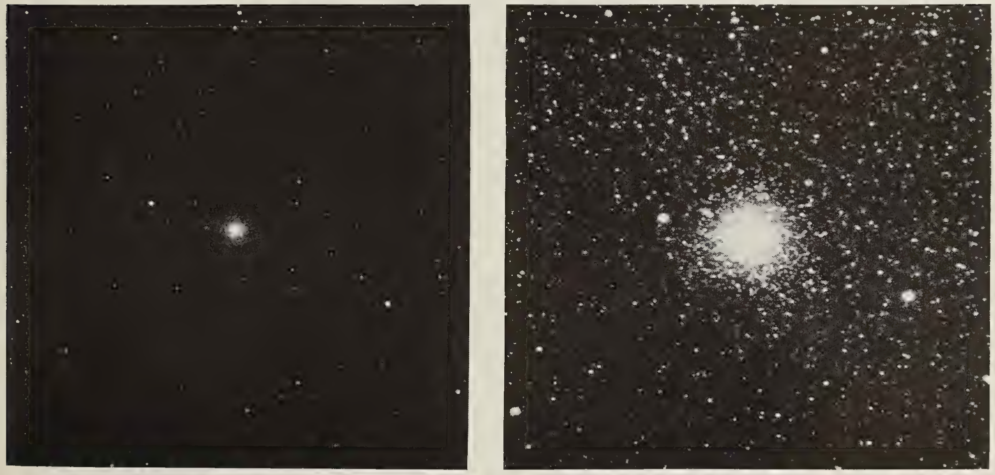


FIG. 1. Colors of B stars and globular clusters in the Milky Way. The larger the circle, the redder the star. Clusters are indicated by barred circles.



Blue

Red

FIG. 2. Blue and red photographs of a globular cluster. Taken by Baade at Mount Wilson Observatory.

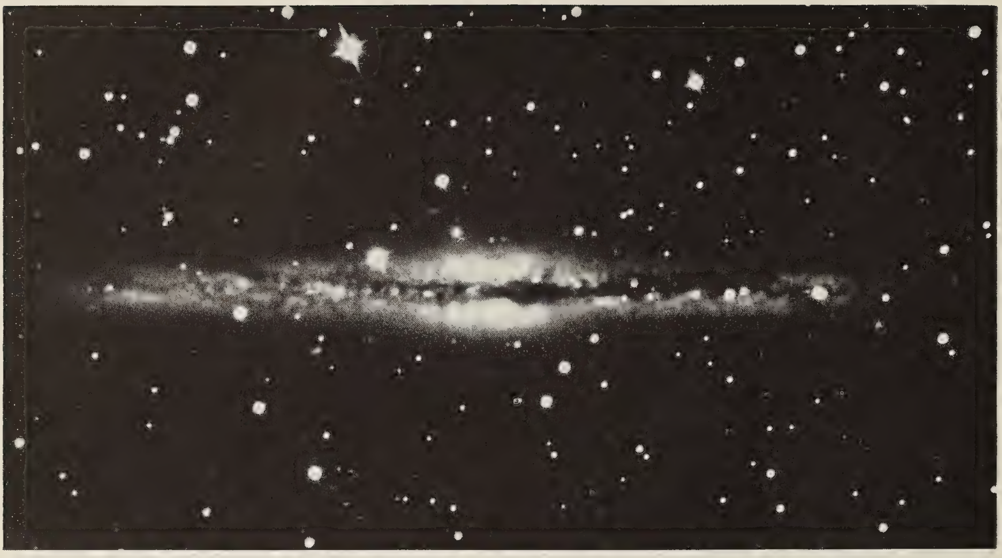


Fig. 1. N. G. C. 891. A spiral nebula or galaxy seen edgewise.

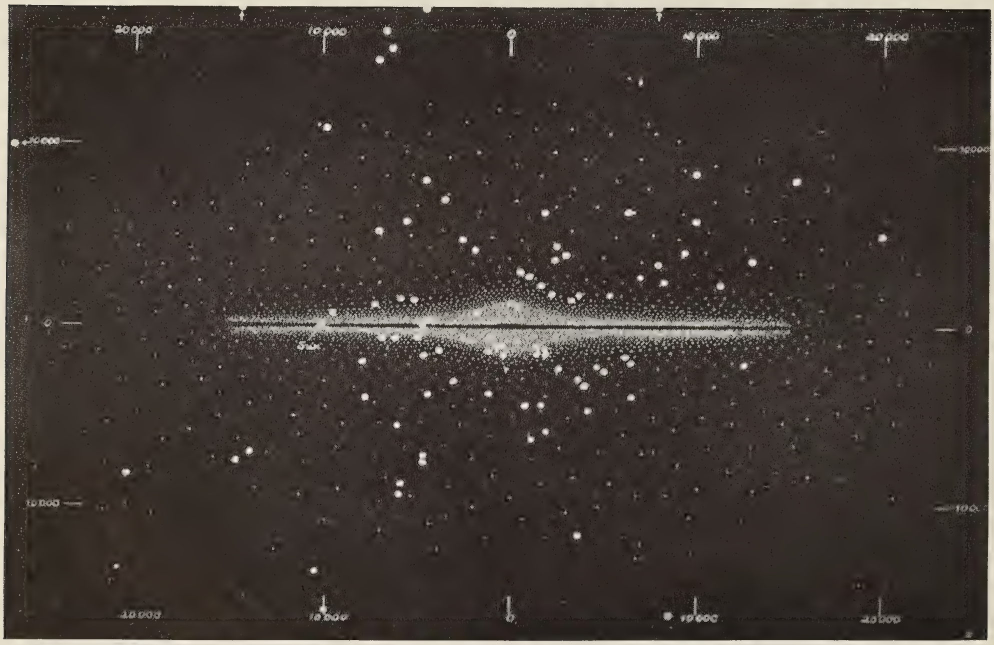

20000
10000
0
10000
20000
10000
0
10000
20000
10000
0
10000
20000

Fig. 2. Schematic diagram of the galaxy by Plaskett. Dimensions are in parsecs; 10,000 parsecs = 33,000 light years.

EXPLORATIONS IN THE REALM OF THE NEBULÆ

EDWIN HUBBLE

Mount Wilson Observatory
The Carnegie Institution of Washington

THE decade centered around 1930 marks an epoch in the large-scale explorations of the universe. Previously, the realm of positive knowledge was roughly coextensive with our own stellar system, called the galactic system or system of the Milky Way. The vast swarm of stars was known to be isolated, but the extent and contents of the surrounding space were largely matters of speculation. During the decade, empirical explorations swept far out into that space which was found to be thinly populated by the extragalactic nebulæ. These nebulæ were identified as independent stellar systems, comparable with our own, and their distances were determined. The results were important because they established the position of our own stellar system in the order of nature, they permitted the comparative study of stellar systems, and they defined an observable region so vast that it may be tentatively considered as a fair sample of the universe. The results, in short, provided an observational approach to cosmology.

THE PROBLEM OF THE DISTANCE SCALE

The situation at the beginning of the decade was somewhat as follows. Distances of many of the nearer stars were known from direct measurement. The study of these stars had led to the identification and the calibration of various criteria of intrinsic luminosity (or candle power). Wherever the criteria could be recognized, the intrinsic luminosities of the stars were known, and, consequently, their apparent faintness indicated their distances. This powerful, indirect method of estimating distances extended the field of empirical investigation far beyond the limits of direct measurement.

The spectrographic criteria of intrinsic luminosity, which represent one of the major contributions of the Mount Wilson Observatory, applied to all stars whose spectra could be recorded with moderate dispersions. Beyond the range of the spectrograph, colors and light variation furnished other criteria. The most precise and reliable of the long-range criteria was (and is) the period-luminosity relation among Cepheid variables. Cepheids are pulsating stars which brighten rapidly and fade slowly, the rate of pulsation varying with the mean size and luminosity. Large, bright stars pulsate slowly, while small, faint stars pulsate rapidly. This relation was discovered by Miss Leavitt at Harvard (1912), was calibrated by Hertzsprung at Leiden (1913), and the calibration was revised in its current form by Shapley at Mount Wilson (1918).

Cepheids are giant stars, ranging from a few hundred times to several thousand times brighter than the sun, and they can be easily recognized and measured even when near the limits of telescopes. Consequently, the Cepheid criterion can be applied out to enormous distances. Shapley, for instance, used the criterion in his classical study of the globular clusters, made at Mount Wilson, which led to the first convincing estimate of the over-all dimensions of the galactic system, and the position of the sun with respect to the center. Eventually, it was the Cepheid criterion which bridged the space between the galactic system and the nearer nebulæ.

At the time, however, the nebulæ were mysterious objects—the only objects in the sky which were not definitely known to be constituent members of the stellar system. They were found in great numbers, everywhere in the sky except along the Milky Way itself. Because of their distribution, they were called "extragalactic" nebulæ, but in this discussion, the adjective is dropped, and the simple term "nebulæ" is employed.

On the photographic plate, the nebular images in general were elliptical, and were presumably the projections of lenticular bodies oriented at random in space. Central nuclei were conspicuous, surrounded by more or less extensive nuclear regions of structureless nebulosity. In most cases, but not in all, two arms emerged from opposite sides of the nuclear region, and wound outward along spiral paths. The spiral arms were frequently resolved into numerous small condensations which were often described as "nebulous" stars.

The nebulæ, it was recognized, might be relatively small bodies, scattered among the stars, or immense bodies distributed through the space which surrounds the stellar system. The crucial question was the order of the distances, and, on this question, the evidence was contradictory. On the one hand, the large radial velocities, measured by Slipher at Flagstaff, suggested that the nebulæ were beyond the gravitational control of the stellar system. On the other hand, the appreciable angular motions in several spirals, measured by van Maanen at Mount Wilson, implied relatively small distances.

Further evidence was furnished by novæ, the only familiar objects that were observed in the nebulæ, but here also the interpretation of the data was controversial. For instance, a nova of the seventh magnitude (S Andromeda, 1885) was known to have appeared in the great spiral in Andromeda, Messier 31. Later, from 1917 onward, numerous faint novæ, averaging about the seventeenth magnitude at maximum, were recorded at Mount Wilson. Thus two classes of novæ were observed in the same nebula, and one class was several thousand times brighter than the other. The distance of the spiral, and its location within or without the stellar system, depended upon which of the two classes was comparable with the familiar galactic novæ.

THE METHOD OF ATTACK

Just before the beginning of the decade, a general review of the problem of nebular distances suggested that the most promising line of attack lay in the possibility of recognizing, within the nebulæ, stars, clusters, or other bodies which were sufficiently well known to permit estimates of their intrinsic luminosities. The project involved an investigation of the contents of nebulæ under the most critical conditions and with the greatest telescopic power available. Consequently, the program was especially suitable for the 100-inch reflector at Mount Wilson.

Two problems were recognized as of basic importance because both were concerned with stars, and stars were the best distance indicators known at the time. One problem was the general search for stars among the condensations in the spiral arms. The immediate question was whether the observed lack of definition was inherent or was a photographic effect introduced by the unresolved nebulous background or by instrumental limitations. The second problem was the detailed and statistical study of novæ (obviously stars) in the hope of determining which, if either, of the two classes was comparable with galactic novæ.

The Solution of the Problem

Distances derived by the Cepheid criterion. The observing program was initiated in 1923, and attention was concentrated on the two most conspicuous spirals in the sky, Messier 31 and Messier 33. Owing primarily to the great power of the 100-inch reflector, the problems were solved quickly, and in an unexpected manner. Under critical conditions, the condensations lost their nebulous appearance on the photographs, and were indistinguishable from normal stellar images. Furthermore, the study of the condensations soon led to the discovery of numerous variables, most of which were readily recognized as Cepheids. These data were decisive. The condensations represented a homogeneous group of apparently stellar objects, and certain of them were definitely identified as giant stars. Consequently, it was reasonable to assume that the condensations in general were giant stars.

The Cepheids, since they exhibited the familiar period-luminosity relation, established reliable zero points for the scales of intrinsic luminosities within the two spirals. A large amount of additional data, previously accumulated, immediately fell into place. The fainter class of novæ in Messier 31 proved to be comparable with galactic novæ, and the bright star of 1885 found its place among the rare "supernovæ." A consistent picture of the condensations emerged; they were the brighter giants and supergiants of two vast stellar systems, comparable with our own system of the Milky Way, but separated from us by nearly a million light years of empty space.

Meanwhile, the investigations had been extended to other conspicuous nebulæ, with similar results. By 1925, a small sample collection of nebulæ was available, whose distances, and, consequently, whose actual dimensions and luminosities were indicated by the Cepheid criterion. Further investigations soon demonstrated that these systems were representative examples, and were conspicuous merely because they happened to be situated in our immediate neighborhood. The smaller, fainter nebulæ were similar systems but more remote.

Actually, the preliminary sample collection, together with the system of the Milky Way, represent the members of a "local group" which is more or less isolated in the general field of nebulæ. Our own membership in such a group is a fortunate accident which permits the detailed study of several stellar systems at unusually close range. This circumstance materially advanced the opening of the realm of the nebulæ to actual exploration. Even today, no Cepheids have been identified with certainty beyond the limits of the local group.

The criterion of brightest stars. The step from the local group to the nearer nebulæ in the general field was made by using the brightest stars in nebulæ as distance indicators. The stellar contents of the various members of the local group, as far as they could be recognized, proved to be remarkably similar to the assemblage of brighter objects in the galactic system. Brightest stars, novæ, globular clusters, Cepheids, and other recognized objects were all in the familiar order of intrinsic luminosity, and any of them, in principle, might have served as a criterion of distance. Actually, the "brightest star," defined as the mean of the three or four brightest individual stars in a system, was adopted as the most useful criterion. The intrinsic luminosities of brightest stars, it was found, varied from nebula to nebula, but the range was small. Since the average luminosity, about 50,000 suns, was much brighter than that of normal Cepheids, the new criterion, although less precise than the old, could be applied throughout a much larger volume of space. Thus it was possible to assemble a large sample collection of nebulæ in the general field, whose distances, dimensions, and luminosities were fairly well known.

The general criterion of total luminosities. The next and most general indicators of distance were evidently the total intrinsic luminosities of the nebulæ themselves. In order to apply this criterion, it was necessary to determine the mean luminosity of nebulæ and the dispersion of individual luminosities around the mean. The sample collection assembled on the basis of brightest stars furnished the information for resolved nebulæ, but other types of nebulæ were known in which stars could not be detected, even in the most conspicuous examples. Consequently, it was necessary to investigate the relative luminosities of the different types.

The problem was first solved by the study of great, compact clusters of nebulæ, in each of which nebulæ of all types were mingled together at a given distance from the observer. Later, it was possible to derive similar data for isolated (non-cluster) nebulæ of all types from the study of red shifts in nebular spectra, which were found to be individually reliable indicators of relative distances. The investigations revealed no appreciable systematic variation in the total luminosities of nebulæ of different types. All nebulæ appeared to be members of a single family, and their luminosities were all of the same general order. Consequently, it was permissible to calibrate numerically the criterion of total luminosity by means of the resolved nebulæ alone.

The results may be stated as follows. The mean photographic absolute magnitude of nebulæ in a given volume of space is about $M_{pg} = -14.2$, with a dispersion of about 0.85. In other words, most of the nebulæ lie between one-half and twice the average luminosity of 85 million suns, although occasional giants and dwarfs are found which differ from the average by a factor of 10. Thus the total luminosities of nebulæ do not furnish precise individual distances but they serve very well as statistical criteria. Where large numbers are concerned—clusters with several hundred members, or extensive surveys—the nebulæ can be treated as though they were equally luminous, and reliable distances can be derived from their apparent faintness.

Once the series of distance criteria was established, it was possible to proceed along two lines. On the one hand, the nebulæ could be studied as stellar systems. They fall naturally into an ordered sequence of types, and a quantitative study of the sequence offers an observational approach to the ultimate problem of nebular evolution. On the other hand, the nebulæ could be used as gigantic beacons, by reference to which the whole of the observable region of space could be surveyed. The study of the general characteristics of this sample of the universe offers an observational approach to cosmology. Preliminary reconnaissances have been made in both fields. Today, the problems are fairly oriented, and significant programs of further investigation can be formulated as more powerful instrumental equipment becomes available.

The Nature of the Nebulæ

The luminosities of nebulæ have already been mentioned. Diameters increase steadily throughout the sequence of types, ranging from an average value of about 1000 light years for the compact globular nebulæ at one end, to about 10,000 light years for the open spirals at the other end. Stars have actually been detected only in the later-type (more open) spirals and in a few irregular nebulæ. Nevertheless, there are strong grounds for assuming that all nebulæ are stellar systems. The upper limit of stellar luminosities presumably increases along the sequence of types, and rises above the observable threshold only in the later section of the sequence. Masses are uncertainly estimated as of the order of 10,000 million suns. Our own stellar system is believed to be a giant, open spiral, more or less comparable in size with the giant intermediate spiral, Messier 31. The small-scale distribution of nebulæ is quite irregular. Nebulæ are found singly, in pairs, triples, groups, clusters, and, possibly, clouds. On the grand scale, however, the irregularities tend to average out, and one large region of space is very much like another. The average distance between neighboring nebulæ is of the order of two million light years.

The Spectra of Nebulæ

The results summarized in the foregoing paragraphs were derived solely from the study of direct photographs. Data of a quite different kind were assembled when the difficult problem of recording nebular spectra was solved. Nebulæ at best are faint objects. They are dominated by semistellar nuclei from which the surface luminosity falls away in all directions to more or less undefined boundaries. The luminosity gradients are steep in the compact globular nebulæ but they diminish steadily through the sequence of types until, in the open spirals, they are almost flat. Thus the nebulæ may be roughly described as ranging from small, concentrated, semistellar objects to large, uniform disks. For the purpose of recording spectra of the latter, the dimensions of telescopes are relatively unimportant; success is measured by the efficiency of the spectrograph, and, in particular, by the speed of the camera behind the prisms. On the other hand, the compact nebulæ tend to behave like stars, and respond to the light-gathering powers of telescopes. Thus it is readily understood that the study of nebular spectra depends both upon the development of telescopes and upon the construction of very fast, low-dispersion spectrographs.

Following the early, experimental work of Scheiner and others, Slipher, in 1912, attacked the problem with an especially designed spectrograph attached to the 24-inch refractor at Flagstaff. By 1925, he had assembled spectra of 41 out of the 45 nebulæ observed up to that time.

The spectra in general were similar to the solar spectrum. The conspicuous pair of calcium lines, H and K, were unmistakable, and familiar lines of hydrogen and iron could be easily recognized. Slipher found that, in all but two or three of the nebulæ, the lines were slightly displaced to the red of their normal positions, and he made the customary assumption that the displacements indicated radial velocities. However, the facts that the velocities were large (they ranged up to 1800 km/sec) and that the motions were almost invariably away from the observer were unexpected features which complicated the interpretation of the data.

The Velocity-Distance Relation or Law of Red Shifts

The first problem was evidently the determination of the motion of the observer (presumably the combined motion of the earth, the solar system, and the stellar system) with respect to the nebulæ. If the nebulæ were drifting more or less at random, their individual motions should tend to cancel out, leaving the reflection of the observer's motion to emerge from the totality of the data. The problem, however, was complicated by the apparent recession of nebulæ in all directions. In order to compensate this systematic effect, Wirtz, in 1918, introduced a constant K-term, a constant red shift of apparent velocity of recession that should be subtracted from all observed velocities before the latter were analyzed. The results, however, were not satisfactory; residual velocities of the nebulæ, after the K-term had been applied and the observer's motion had been removed, remained conspicuously systematic.

The failure of a constant K-term led naturally to the trial of a variable K-term, a red shift, or apparent velocity of recession which increased with distance. Solutions for the observer's motion, incorporating the new assumption, were attempted by Wirtz and others in 1924–1925 but, in the absence of reliable criteria of distances, the results were again unsatisfactory. Finally, in 1929, after the criterion of brightest stars had been established, the problem was solved at Mount Wilson. The observer's motion was largely accounted for by the rotation of the sun around the center of the stellar system, and the K-term proved to be directly proportional to distance. In addition to their individual motions (which were more or less at random and averaged about 200 km/sec), the nebulæ exhibited a systematic velocity of recession (real or apparent) which increased at the rate of about 500 km/sec per million parsecs of distance (or about 100 miles/sec per million light years).

The so-called "velocity-distance" relation was immediately recognized both as a remarkable property of the universe and as a valuable criterion of distance. Consequently, it was imperative that the relation be investigated over the greatest possible range in distance and calibrated with the greatest possible precision. Slipher's list of velocities offered a fair sample of the brighter, near-by nebulæ; it was now necessary to observe spectra of the fainter, more distant systems.

At this point began the joint investigations of nebulæ at Mount Wilson, which rapidly developed into effective group cooperation.[1] Humason, using improved spectrographs with the large reflectors, undertook the exacting problem of assembling spectra of faint nebulæ, while Hubble determined the distances. Attention was first concentrated on the brighter nebulæ in the great clusters whose relative distances were indicated by the apparent faintness of their numerous members. The research was exciting for, with each improvement in equipment, fainter nebulæ were observed, and the spectral lines marched steadily along the spectrum toward the red.

[1] Cooperation is an important and distinctive feature of nebular research at Mount Wilson. The group includes Baade, Hubble, Humason, Minkowski, Stebbins, and Whitford, and (until his recent death) Sinclair Smith, working in intimate association with Bowen, Tolman, and Zwicky at the California Institute of Technology. Among these individuals, with their widely various interests and special qualifications, there is a marked tendency toward combining resources in particular investigations, and interpreting the results in the light of constructive criticism of the group as a whole. A current example of the procedure is the study of supernovæ in nebulæ by Baade, Minkowski, and Zwicky, each concentrating on a different aspect of the problem and submitting the results for general discussion before their final formulation.

The group cooperation extends through an even wider field because of the close and friendly relations with the director and staff of the neighboring Lick Observatory at Mount Hamilton. Data and ideas are exchanged, and programs are adjusted to avoid undue overlapping, with that confidence in the integrity of investigators which should always characterize scientific research.

With data for clusters now ranging over a distance of the order of 250 million light years, where the red shifts correspond to velocities of the order of 25 thousand miles per second (one-seventh the velocity of light itself), the sensibly linear form of the relation is established out to the effective limits of existing equipment.

Meanwhile, the spectra of many field nebulæ, including those in which stars can be seen, have also been recorded. The resolved nebulæ, because of their known distances, served to calibrate the relation—the latest figure is 526 km/sec per million parsecs, or 101 miles/sec per million light years. The total assemblage, since it included all types, permitted a comparative study of different types. When distances indicated by red shifts were used, the mean, intrinsic luminosities of nebulæ were found to be fairly uniform along the sequence of types, and all nebulæ, whether in clusters or in the general field, could be considered as comparable systems.

The Significance of Red Shifts

The interpretation of red shifts is not yet definitely established. The shifts represent a loss of energy in the light which reaches the earth. If the loss occurs in the nebulæ, the shifts almost certainly are velocity shifts, indicating a general recession of the nebulæ at velocities which are directly proportional to distances. On the other hand, the loss might occur (by some unknown reaction) in the intervening space. If the contents of internebular space were moderately homogeneous, the loss would necessarily be proportional to the distance traveled. In this case, the system of the nebulæ would be sensibly stationary.

The choice between the two working hypotheses is of fundamental significance in cosmological theory, and, in principle, the question can be determined empirically. A receding nebula should appear fainter than a stationary nebula at the same momentary distance. The dimming effect, or "recession factor," is negligible for ordinary speeds but increases with velocity and becomes appreciable for the large red shifts that are found, or might be found, near the limits of existing telescopes. Thus, for a given red shift, a given apparent faintness might represent either of two distances, depending upon whether the recession factor is applied or omitted. The test cannot be made directly because the only available criterion of great distance is the apparent faintness itself. Consequently, it is necessary to explore the observable region of space on the basis of the alternative scales of distance in the hope that the wrong scale may lead to discrepancies which will be recognized.

The Observational Approach to Cosmology

In a practical sense the program consists in the precise formulation of the law of red shifts (as the velocity-distance relation is properly called) and of the law of nebular distribution. The investigations have been carried out at Mount Wilson, where red shifts, as previously mentioned, have been studied out to distances of the order of 250 million light years, and nebular counts, to a distance of the order of 400 million light years. In both cases, the penetration represents about the limits for reliable measurement with the 100-inch reflector. The results, however, are suggestive rather than definitive.

When recession factors are ignored, the law of red shifts is sensibly linear, and the large-scale distribution is sensibly uniform over the complete range of the observations. The observable region then appears to be thoroughly homogeneous, and is too small a sample to indicate the nature of the universe at large. Nevertheless, the uncertainties of the data are sufficient to permit a consistent, although somewhat forced, interpretation when recession factors are included. These factors distort the previously used scale of distances in such a way that the law of red shifts is no longer linear (red shifts increase with distance at an accelerated rate) and the apparent distribution is no longer uniform (the density increases outward). The sample then becomes significant, and seems to identify a particular type of universe among the long array of possible types offered by the current theories of relativistic cosmology. The type in question is decidedly strange—a finite universe that is small and quite young. Thus a choice seems to be presented between an indefinitely large, stationary universe and a small, finite, expanding universe. The observational data favor the former, but do not rule out the latter alternative.

At this point the explorations are stopped by the limitations of instrumental equipment. The conquest of the realm of the nebulæ has been the achievement of a great telescope. It was the 100-inch reflector which furnished the essential clue, namely, the distance scale. The rapid development in the new field represents the exploitation of this single clue out to the very limits of the 100-inch. The observable region now defined by the great telescope is a sample of the universe so large that it justifies the first tentative steps in an empirical approach to cosmology. The problems have been formulated, methods devised and tested, solutions restricted to a fairly narrow range. The way thus seems to be open for definitive investigations when still greater instrumental power becomes available.

SOLAR AND TERRESTRIAL RELATIONSHIPS

SETH B. NICHOLSON
Mount Wilson Observatory
The Carnegie Institution of Washington

THREE years ago, in 1935, Dr. J. H. Dillinger of the National Bureau of Standards called attention to a new terrestrial phenomenon which he suspected might be of solar origin. The story of the investigation of this phenomenon, with its far-reaching results in seemingly widely separated fields, illustrates well the value of scientific cooperation; and its narration here seems especially fitting, since some of the leading characters in it were members of the Carnegie Institution of Washington. Since the story is told by a member of the solar department of the Mount Wilson Observatory, the setting and action will be mainly astronomical, with relatively little mention of the equally important and fascinating chapters relating to other scientific fields.

For at least a hundred and fifty years the possibility that changes in the sun might produce significant effects on the earth has been considered and enthusiastically investigated. As early as 1800 Sir William Herschel compared the price of wheat with the number of spots on the sun, thinking that some relationship might exist. He was primarily interested in whether the presence of spots on the sun had any effect on the weather and thought that the price of wheat was a better index of weather conditions over a long period of time than the records of meteorologists. He found a slight apparent correlation but realized the limitations of his data and did not consider that any positive relationship had been established. He merely called attention to the possibility, warned anyone from giving the matter more consideration than the observations showed it was worth, and hoped that more accurate data in the future would settle the problem. Since then thousands have attempted to connect various terrestrial phenomena directly with sunspots or to show that some indirect relationship exists. A great variety of variables have been considered—rainfall, business conditions, the quality of wine, the fecundity of rabbits, the number of automobiles manufactured, the severity of influenza—these and many others equally fantastic have been investigated and in most cases have been found by their champions to be related to changes in the number of sunspots.

Like Herschel, they usually realize the limitations of their data and say that a longer series will be necessary to establish the relationship. The investigator today, with much more reliable data than those used by Herschel and with a record nearly a hundred and fifty years longer still hopes that more data will solve the problem. In many cases longer series of observations have weakened rather than strengthened the correlation, thus showing that the suspected relationship was only a temporary coincidence. Such failures seldom receive much publicity, and the idea that a correlation exists may persist long after additional observations have discredited it.

Of the many variables tested, the mean temperatures of tropical stations on the earth may be cited as an example. The mean annual temperatures of six tropical stations were obtained from the "World Weather Records," volume 70 of the *Smithsonian Miscellaneous Collections*, for a period of fifty-eight years and compared with the sunspot data for the corresponding years. The mean temperature was found to have been about half a degree lower at sunspot maximum than at sunspot minimum, and since other stations give essentially the same result, the difference is probably real, although we should follow Sir William Herschel's example and warn the weather prophets not to use it for more than it is worth. The data show that if we make no allowance for sunspots, we may predict that the mean annual temperature of the six stations will be equal to its average for the past fifty-eight years, giving odds of one to one that the prediction will not be in error by more than three-tenths of a degree. On the other hand, if we take into account the phase of the solar cycle the probable error of the prediction will be improved by only five-hundredths of a degree.

Many of the correlations which have been investigated, such as the possibility that sunspots indicate some change in solar radiation which in turn changes the air we breathe and the food we eat and thus influences our physiological or psychological behavior, have little more scientific basis than the supposition that our nervous system reacts to changes in the earth's magnetic field or that our lives are regulated by the conjunctions of the planets.

There is no evidence that an individual sunspot has any physical effect whatever on the earth. Sunspots radiate only half as much heat as the normal solar surface, and if the spotted area were sufficiently great, the reduction in solar radiation might affect the earth or could at least be measured; but the largest spotted area ever observed would not reduce the solar radiation by as much as 0.5 per cent, a quantity too small to be distinguished from the greater variations due to changing absorption in the earth's atmosphere. All solar features—sunspots, prominences, coronal streamers, and faculæ—vary in a cycle of about 11 years, and it is quite possible that these changes in the general activity may be accompanied by a slight change in the average temperature of the sun's radiating surface. A change of 10°C. would produce a variation in heat greater than that due directly to changes in the spotted area.

There is, however, little observational evidence of such a change. Direct measurements of the solar radiation made by the Astrophysical Observatory of the Smithsonian Institution since 1920 indicate fluctuations of very short period but show no cyclic variations with the sunspots. The mean value of the solar radiation for 1926, 1927, 1928, and 1929—years of great solar activity—was 1.939 calories per square centimeter per minute, while that for the years of low solar activity, 1922, 1923, 1924, 1932, 1933, 1934, was also 1.939. The average for all the years from 1920 to 1934 was 1.941.

It is, of course, possible that significant changes may take place in the quality of sunlight without any appreciable change in the quantity of radiation emitted. Ultraviolet light plays an important part in life processes on the earth, and changes in the percentage of ultraviolet light would almost certainly have many important consequences. Measurements of the ratio of ultraviolet to green light have been made at the Mount Wilson Observatory by Dr. Edison Pettit since 1924. These measures concern the radiation between wave lengths 3100 A and 3300 A transmitted by a silver film and that between 4500 A and 5500 A transmitted by gold film. Ultraviolet light of very short wave length is not transmitted by the earth's atmosphere and is therefore not subject to direct measurement at the earth's surface. Dr. Pettit's measures show large changes in the ratio of ultraviolet to green light, and for the first few years these changes seemed to be related to changes in the spottedness of the sun; but the later measurements do not show a striking correlation, especially in relation to the 11-year solar cycle.

Although individual sunspots do not directly affect the earth, they are visible manifestations of unusual activity on the sun and it is well established that certain terrestrial phenomena, notably auroral displays and rapid fluctuations in the earth's magnetic field, known as magnetic storms, tend to occur more frequently when the presence of many sunspots indicates that the sun is unusually active. The greatest magnetic storms and the most brilliant auroral displays nearly always occur when an exceptionally large and active group of spots is present; but one or two notable exceptions have occurred, and there is other evidence that the cause of magnetic storms emanates from areas on the sun in which sunspots may appear but that the spots themselves are not absolutely essential to the occurrence of the storms. On the other hand, quite large groups of spots are sometimes carried across the solar disk by the sun's rotation without producing either magnetic storms or auroræ.

One of the characteristics of the groups associated with magnetic disturbances is the fact that bright eruptions of the chromosphere seemed to occur in them more often than in groups not associated with magnetic activity. These chromospheric eruptions are comparatively short-lived phenomena, lasting from 20 to 60 minutes, although some have persisted for a few hours. In a few cases very intense eruptions of this sort were followed some hours later by intense magnetic storms and it seemed possible that the eruptions were actually the cause of the storms. In the daily routine of solar observing only a very limited number of these eruptions had been recorded, and it was obviously desirable that their frequency and intensity should be studied in detail.

In 1931, Dr. Hale, realizing the importance of a more complete record of these eruptions, suggested that solar observers equipped with spectrohelioscopes cooperate in an attempt to make a complete continuous record of solar activity. At that time, twenty-three spectrohelioscopes and cœlostat telescopes had been built or ordered for stations well distributed in longitude around the world.

The spectrohelioscope is well adapted to this problem. The few bright chromospheric eruptions that had been recorded were conspicuous objects on spectroheliograms made with the light of the

H and K lines of calcium or with the red line of hydrogen, and the spectrohelioscope had been specially designed to use the red hydrogen line. Dr. Hale's remarks in a paper on "Solar Eruptions and their Apparent Terrestrial Effects"[1] express so well the state of the problem that they are quoted here.

> It is important that all signs of eruption on the disk, especially those observed near the central meridian of the sun, be followed and carefully analyzed. Many small eruptions may break out suddenly and subside within a short time, but others, of similar appearance in their initial stages, may develop into vast outbursts, of intense brilliancy and of great importance from a geophysical standpoint. If, as the available evidence so strongly suggests, rapid changes in intense flocculi near the central meridian of the sun are usually followed by terrestrial disturbances, it is highly desirable that all the significant facts, such as their times of appearance, maximum development, and disappearance, and their approximate heliographic position, form, area, motions, and intensity at successive stages, be learned as soon as possible. The various theories hitherto proposed, especially those which attribute to radiation pressure the flight of erupted gases from the sun to the earth, can then be thoroughly tested, thus greatly facilitating other studies of solar and terrestrial relationship. Existing data are very incomplete, as might be expected because of the small number of observations made daily and the distribution of properly equipped observers in longitude. The spectrohelioscopes now available are more numerous, better distributed in longitude, and much quicker in action than the few spectroheliographs in daily use. They also have the advantage of permitting the instant change in wave length of the light under observation, thus enabling effects often missed with the spectroheliograph to be detected and analyzed.

Dr. Hale hoped that, if a complete record of bright eruptions were available, we might establish, among other things, the time lag between the solar outburst and the subsequent magnetic disturbance by identifying a series of eruptions with a series of disturbances spaced at the same time intervals.

At the meeting of the International Astronomical Union at Cambridge, Massachusetts, in the fall of 1932, the organization of spectroheliograph observers was taken over by Commission 11 under the presidency of Dr. L. d'Azambuja. Observing times were assigned to each station and arrangements were made to collect the observations at Meudon and to prepare them there under d'Azambuja's direction for publication in the *Bulletin for Character Figures of Solar Phenomena.* This bulletin, also a project of the International Astronomical Union, is published at Zürich under the direction of Professor W. Brunner, who is president of Commission 10 of the Union.

---

[1] *Mt. Wilson Contr.*, No. 425; *Astrophys. Jour.*, vol. 73, p. 379, 1931.

Solar activity was at a minimum in 1932 and 1933, and the observers had little to report until activity in the new cycle had commenced. Many of the observers were inexperienced in this type of observing and were somewhat at a loss to know just what sort of phenomena to report. Until they had actually seen a really bright chromospheric eruption, they had no idea of the intensity that such an eruption might attain. Interest in this program was greatly stimulated when Dr. Dillinger called attention to the occasional complete fading of high-frequency radio signals, so sudden and so complete that radio operators thought the power had been turned off or that something had broken in their receiving sets. Frantic investigations failed to find anything wrong, and the operators were more mystified than ever when, after thirty minutes or an hour of silence, messages again came through uninterruptedly. After a few repetitions of this sort of experience, a check-up revealed that the fading was general wherever the sun was shining but that no such effect was experienced on the dark hemisphere of the earth. It was then, in 1935, that Dr. Dillinger wrote to the Mount Wilson Observatory, inquiring whether any unusual solar activity had been observed at the times of these fade-outs. Fortunately, either hydrogen or calcium spectroheliograms were available at the exact times of all but one of the fade-outs which he reported. In every case for which photographs were available a bright chromospheric eruption was found to have been in progress.

Since 1935 a much more nearly continuous watch of the sun has been maintained than ever before. In May 1936, an automatic camera was attached to the spectroheliograph on Mount Wilson, which photographs a 2-centimeter image of the sun on 35-mm motion-picture film. The interval between exposures is ordinarily three or four minutes, the actual exposure being a little less than one minute. This instrument is operated continuously whenever the 60-foot tower telescope is not otherwise in use. As a result, a fairly complete record of the sun's appearance is available for all the clear days on Mount Wilson since May 1936.

Dr. R. S. Richardson of the Mount Wilson Observatory staff is in charge of this program and makes quarterly reports to Meudon of the times and intensities of all the eruptions observed visually or recorded on the photographs. Before the automatic camera was installed, he took many special spectroheliograms in an effort to record and obtain the intensities of as many eruptions as possible. He has also catalogued all the eruptions recorded on the routine daily spectroheliograms made at Mount Wilson since 1917. A special spectrograph, with which the entire visible spectrum can be photographed with one exposure, has been installed, and Dr. Richardson has obtained spectra of a few bright eruptions with it. This necessitates a continuous watch of the sun with the spectrohelioscope in order that the eruption may be detected as soon as it takes place.

Two of the Hale spectrohelioscopes were loaned to the Department of Terrestrial Magnetism of the Carnegie Institution for use at their observing stations in Peru and in Australia. On April 8, 1936, the observers at Huancayo, Peru, were especially fortunate in obtaining at this one station a very complete record of a bright chromospheric eruption and its various terrestrial effects. The beginning of the eruption was observed visually with the spectrohelioscope; and the records of changes in terrestrial magnetism, in earth currents, and in the strength of reflected high-frequency radio signals all show disturbances beginning at the same time that the gases on the sun had been seen to erupt. Photographic observations of this eruption made at Mount Wilson confirmed the observations at Huancayo, and the Mount Wilson magnetic records supplied additional data necessary for a study of the geographical distribution of the magnetic effects.

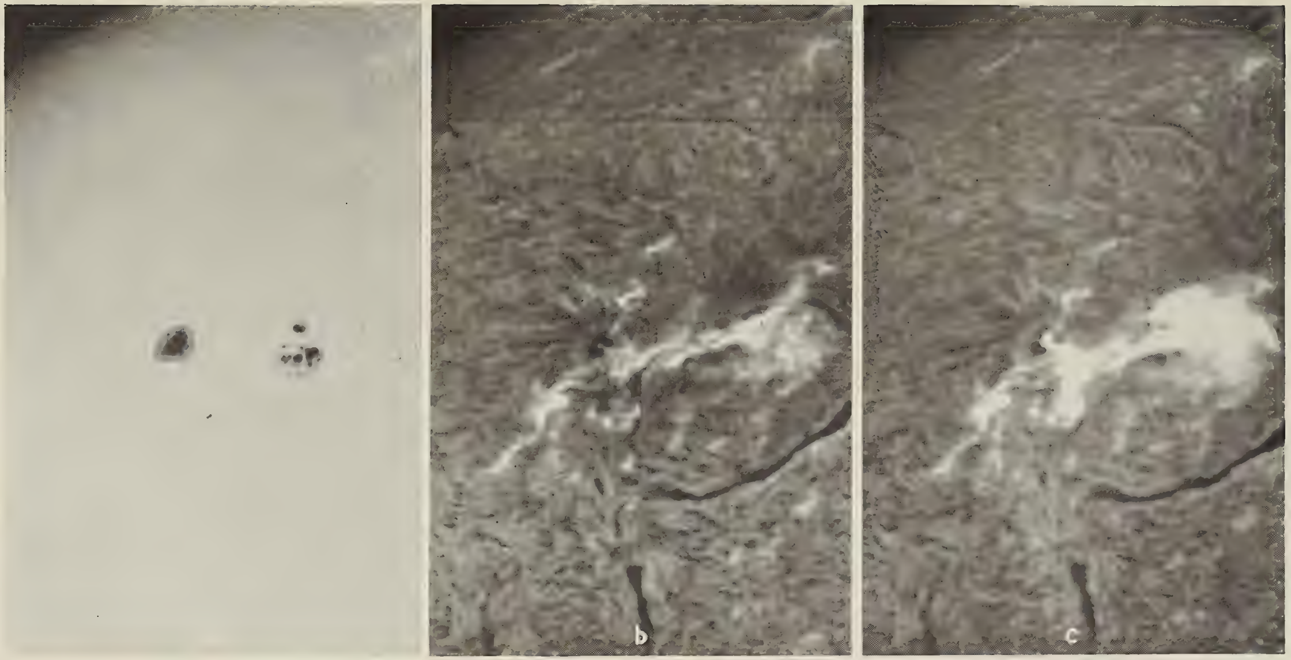
Several hundred bright chromospheric eruptions have now been reported by the chain of cooperating solar observers, and much has been learned in the past three years about these solar phenomena and their terrestrial relationships. Such eruptions with their direct simultaneous effects on the earth are the first and only evidence that has so far been obtained of definite terrestrial changes produced by specific solar phenomena.

When the sun is seen through a spectrohelioscope which is adjusted to transmit only the light of the red hydrogen line, its appearance is very different from that of a direct view. The light hydrogen gas lies high above the heavier gases of the solar atmosphere, and in some places large masses of it rise to still higher levels, forming what are called prominences. These may be seen in profile at the edge of the sun, extending far above the average level of the hydrogen atmosphere, called the chromosphere. There they appear bright in contrast with the darker background of the sky, but are intrinsically fainter than the chromosphere. On the disk of the sun they absorb more solar radiation than the thinner chromosphere and consequently appear darker than the rest of the disk.

Huge spiral vortices are often found in the hydrogen atmosphere over sunspots, covering an area much larger than the spots themselves. Bright hydrogen clouds also appear in the neighborhood of spots and sometimes even over areas where no spots are visible. They are comparatively stable and their general appearance changes very little from day to day. These bright hydrogen clouds are much less intense than the very bright eruptions which burst forth suddenly, reach their maximum brightness in a few minutes, and then slowly subside. The amazing brilliance of the chromospheric eruptions and the startling rapidity with which they develop make them one of the most spectacular of solar phenomena. They usually occur in the neighborhood of active sunspots, although there is no apparent connection between them and the spots. They extend high up into the solar atmosphere and when seen at the edge of the sun look like brilliant jets or tongues of flame extending outward from the edge of the sun like ordinary prominences only much brighter and not as high as most prominences. When an eruption is at its maximum size and brightness, it completely blocks out any sunspots or other details of the solar surface over which it lies, but when it has faded out the details are seen again apparently unchanged by the tremendous activity that has been going on far above them.

The spectroscope shows that the bright eruptions are composed primarily of the same gases that are found in the chromosphere and in ordinary prominences, namely, hydrogen, helium, and ionized calcium. All their light comes from the radiations characteristic of these gases and although they appear so bright when seen in the spectrohelioscope, with all other light eliminated, they are completely invisible when the sun is viewed directly or photographed with ordinary light. Only one eruption bright enough to be seen without the aid of a spectroscope has ever been recorded; this one was seen by two observers in 1859.

[NICHOLSON] PLATE 1



(*a*) A direct photograph of a sunspot group; (*b*) the same region photographed with the red light of hydrogen only ten minutes before the bright eruption (*c*) was at its maximum intensity.

These phenomena should not be confused with "eruptive prominences," which are entirely different in nature. Ordinary prominences are much more stable than chromospheric eruptions and are usually many times higher and much less brilliant.

Although the gases within the ordinary prominences are in continual motion, the general form of the prominence may be maintained for several days. Occasionally, however, such a prominence will erupt, not suddenly like a chromospheric eruption but in a much more leisurely fashion. The eruption of a prominence is not accompanied by an increase in brightness, and the great velocities of recession from the sun which they sometimes attain are reached by successive stages. Such prominences rise thousands of times higher than the chromospheric eruptions and their velocities, although acquired more slowly, are often much greater than those observed in a chromospheric eruption.

Although bright chromospheric eruptions usually occur over or near spot groups, a few of great intensity have occurred in regions where no spot could be seen. The outbursts are usually between the members of a spot group, but when one does occur near a large single spot it tends to avoid the region directly over the umbra, and appears close to one side of the penumbra. A particular area is often intermittently active, as many as six distinct eruptions having been observed in the same region within a period of nine hours. A most remarkable feature of chromospheric eruptions is that they sometimes occur simultaneously over widely separated spot groups. Such occurrences are too frequent to be chance coincidences, and there are instances when as many as three, and in one case four, widely separated eruptions occurred at the same time. The visible radiation from bright eruptions, although intense, cannot account for the observed terrestrial results, a circumstance which indicates that an immense increase in the invisible ultraviolet light must accompany the eruptions.

Chromospheric eruptions vary greatly in apparent brightness and are grouped by the observers into classes 1, 2, and 3, the brightest being of class 3. About 70 per cent of the eruptions reported are of intensity 1, 26 per cent of intensity 2, and 4 per cent of intensity 3. In 1936 and 1937 an average of 2 eruptions per month of intensity 3 were reported to Meudon and 14 per month of intensity 2.

Radio fade-outs have not been observed to coincide with all these eruptions; and although the percentage is higher for the more intense eruptions, a few of intensity 3 have been reported with no accompanying terrestrial effects. When terrestrial effects are observed, their coincidence in time with the eruption is so exact that no doubt of a direct relationship can exist. Generally the two phenomena take place at exactly the same time within the accuracy of the determination. Obviously the solar observations are not continuous like the magnetic records and some of the radio records, so that when an eruption is first observed there may be an uncertainty of a minute or two in the time of its beginning. There is no record of a fade-out having preceded the eruption with which it is obviously associated. There are a few instances, however, when eruptions were visible a few minutes before the terrestrial phenomena began, and in one well-established case the lag was as much as twelve minutes.

As has been stated, the terrestrial effects are much more pronounced when the sun is nearly overhead at the time of the eruption, and many failures to observe such effects may be due to the geographical location of the radio stations or the magnetic observatories. Not all such failures, however, are due to lack of observations. Some eruptions, apparently as intense as others, do not produce their characteristic terrestrial results. Fade-outs have been produced by eruptions at all distances from the center of the solar disk and apparently do not depend on the location of the eruption on the sun. It may well be that the intensity of the visual light from an eruption is not proportional to the ultraviolet light radiated by it. The critical condition may be that of level in the solar atmosphere, the ultraviolet radiation from a low-lying eruption being absorbed in the sun's atmosphere while the visual light escapes. It is probable that the terrestrial effects are produced by the absorption of monochromatic ultraviolet light from hydrogen or helium by terrestrial gases which absorb only those particular wave lengths. If this is the case, the radial velocity of the emitted gases may affect their efficiency in ionizing the earth's atmosphere. These problems are far from settled, but observations of the spectra of eruptions and the continued study of the nature of their effects on the ionosphere will undoubtedly increase our understanding of both phenomena.

The direct magnetic effect of a solar eruption on the earth is not at all like a magnetic storm. Magnetic storms begin simultaneously all over the earth and are more intense in the polar regions than at the equator. The magnetic effect of a chromospheric eruption occurs where the sun is well above the horizon and is more pronounced at the equator than toward the poles. There are distinct diurnal changes in the earth's magnetic field which are obviously directly connected with sunlight. These changes take place regularly and are most conspicuous on magnetically calm days. The changes in the earth's magnetic field produced by solar eruptions are augmentations of the regular daily changes at all places where they are observed. When an eruption occurs, the behavior is as if two suns were suddenly in the sky instead of one.

The regular diurnal changes in the earth's magnetic field are most reasonably explained by electric currents flowing in the upper atmosphere of the earth, but the level of these currents had not been determined until the effects of eruptions on radio reception and on the earth's magnetic field were compared. From the time required for a reflected radio signal to return to the earth, the height of the layer from which it is reflected can be determined, and in this way it has been found that the principal effects of solar eruptions are not over 60 miles from the surface of the earth and that the electric currents which produce the diurnal changes in the earth's magnetism and which are increased during a solar eruption must lie below that level.

Since chromospheric eruptions are more frequent in large, active spot groups of the sort that are most often associated with magnetic storms, we may expect a statistical correlation between eruptions and magnetic storms. There is some evidence that the relationship is more direct; but if so there are marked differences between the relationship of eruptions to magnetic storms and their characteristic direct effects. If eruptions play any part in the production of magnetic storms, apparently there must be a lag in the occurrence of the storm of several hours, usually between 20 and 40; and eruptions near the center of the solar disk would seem to be somewhat more effective in producing these storms. Such were the relationships about which Dr. Hale hoped to learn when he inaugurated the program for cooperative solar observing. Important data are being accumulated which will undoubtedly clarify these relationships, but they are less direct than the newly discovered phenomena and more observations will be necessary for the solution of their problems.

The understanding of the unexpected phenomenon of sudden radio fade-outs was reached so quickly because of the solar photographs accumulated in the daily routine observing, and especially by the cooperation of many observers in different fields. Fortunately for the solar side of the problem, the organization of observers was complete and ready to function. It is to be hoped that the data thus accumulated will be as valuable in solving the problem for which the organization was developed as they have been in the solution of the new problem which arose so unexpectedly.

KNOWLEDGE OF THE PLANETS IN 1938

THEODORE DUNHAM, JR.

Mount Wilson Observatory
The Carnegie Institution of Washington

Reflected sunlight is our only source of information about the physical constitution of the planets, and so our understanding of them must depend on the methods that can be invented for studying their radiation as it comes to us, and on our ingenuity as interpreters of the resulting data.

Early naked-eye observations of the brightness and position of the planets were superseded by measures of high precision after the telescope was invented. This instrument made possible an examination of surface detail, to which photography was added as a powerful tool. Thermocouples later revealed the temperatures of the planets. The spectrograph has been in use for many years, but its possibilities for studying physical conditions seem to be far from exhausted.

The usual plan, in discussing physical conditions in the solar system, is to describe all that is known about each of the planets, one at a time. But in the present summary this procedure will be reversed, so as to emphasize more particularly the various methods which have been employed in planetary research, and to point out what information has resulted, and may still be expected to result, from each.

Naked-Eye Observations

The eye without a telescope can provide data of two kinds only: brightness and position, and the change of both with time. Incorrectly interpreted, these data for the planets were the basis for the Ptolemaic system. Correctly interpreted, they supplied Aristarchus and, later, Copernicus with a more reasonable picture of our solar system. The invention of the quadrant and the astrolabe greatly increased the accuracy of observed planetary positions. These instruments, in the hands of an observer like Tycho Brahe, yielded observations adequate for the important work of Kepler and of Newton. When at last the law of gravitation was clearly enunciated, it was only necessary to compare the new theory with existing observations in order to establish its accuracy throughout the solar system.

The Telescope

Modern astronomy begins with the invention of the telescope. Before turning to its more dramatic revelations of surface detail on the planets, it is worth while to consider for a moment the revolution which the telescope made in the accuracy with which planetary positions could be determined. The transit circle, originated by Römer in the seventeenth century, produced measures which revealed the relative dimensions and the eccentricities of planetary orbits with a new order of accuracy.

In spite of this precision in the determination of relative dimensions, the absolute scale of the whole solar system remained very uncertain until late in the seventeenth century. All methods for finding the scale, except those based on spectroscopic radial velocities, depend on measuring the change in direction of another member of the system when seen from two positions on the earth. The angular difference in direction at the ends of this short base line is much too small to be measured by the eye, even with the best mechanical aids. The telescope, however, made it possible for Richer and Cassini to observe Mars with considerable accuracy from two widely separated stations in 1672, and the reduction of their observations gave the first reliable determination of the scale of the solar system. They were in error by only 8 per cent. Measurements of the solar parallax have steadily improved, especially through observations on minor planets, such as Eros, whose orbits are well known and which approach the earth even more closely than does Mars.

Our knowledge of the masses of the planets depends also on the telescope. In the case of a planet without a satellite, such as Mercury, Venus, or Pluto, the only means of finding its mass is to detect perturbations which its gravitational pull produces in planets in neighboring orbits. The minute departures from perfectly elliptical paths and the minute distances by which these neighbors are drawn ahead of or behind their scheduled positions can only be measured with telescopic aid. It was through a careful study at Mount Wilson of the attraction of Pluto on Neptune that Nicholson and Mayall showed that the mass of the former is not very different from the mass of the earth. At a much earlier date, Le Verrier and Adams used the inverse method when they predicted the existence and position of Neptune to explain the outstanding perturbations of Uranus. In the same way Lowell and W. H. Pickering were led to their belief in a trans-Neptunian planet.

A more exact determination of the mass of a planet is possible if it has one or more satellites, as in the case of Mars, Jupiter, Saturn, Uranus, and Neptune. It is then only necessary to measure carefully the apparent diameters of the orbits of the satellites and their periods of revolution. The solar parallax is required to give in kilometers the scale of the satellite's orbit.

There are two peculiarities among the outermost members of the solar system. One is the retrograde motion of Triton, the only satellite of Neptune, and the other is the fact that the orbit of Pluto is so eccentric that it intersects the orbit of Neptune. Lyttleton, working at Princeton, has shown that it is not impossible that Neptune once had two satellites revolving in the same direction as other members of the solar system, and that a close encounter between them occurred, as a result of which Triton had its direction of rotation reversed while the other satellite was ejected from the Neptunian system and took its place as an independent planet in the solar system, and is now known as Pluto.

The mean density of each planet can obviously be calculated from its apparent diameter, its distance, and its mass. But the interpretation of telescopic measures of the shape of a planet goes one step farther. It can readily be shown that a planet of given mass, radius, and period of revolution will be flattened at the poles to a degree which depends on the extent to which its total mass is concentrated toward the center, and so measures of ellipticity reveal the amount of this internal concentration. The flattening may also be computed from the orbits of the satellites by measuring the rate of revolution of the nodes and of the line of apsides. Both methods agree in showing that the density of Mars must be nearly uniform, that the earth must be considerably denser at the center than it is close to the surface, and that in the case of Jupiter and

Saturn the concentration of matter in the inner regions must be very great indeed.

Knowledge of the mass and diameter of a planet reveals, moreover, the force of gravity acting at its surface, and this permits important deductions to be made as to the type of atmosphere which may be expected on each of the planets. It is more than likely that the planets were once part of the sun and that they were torn away at a cosmic encounter with a passing star. If so, their original endowment of matter must have had the same chemical composition as the outer layers of the sun. These outer layers can be analyzed with the spectrograph, and are known to contain far more hydrogen, nitrogen, and carbon than does the earth, and a very similar amount of oxygen and the metals. What would happen during the cooling of one of the planets would depend almost entirely on its size and on its rate of cooling.

The larger planets, with their high surface gravity, would retain everything, including the vast excess of hydrogen which exists on the sun. As cooling proceeded the hydrogen would probably combine with the carbon, nitrogen, and oxygen to form compounds. Planets of medium size, such as the earth, Venus, and Mars, would have lost a large fraction of their hydrogen atoms while in their original incandescent condition, because the velocity of thermal agitation of the hydrogen atoms would have exceeded the limiting velocity of escape set by the force of gravity on these planets. A part of the carbon and nitrogen and some of the oxygen has probably gone too, but all heavier atoms would have been retained. Mercury might have held the heavier gases if it were not so hot on account of its proximity to the sun. But the moon, all other satellites, and all the minor planets have so little surface gravity that molecules of every gas would fly away and leave them without any atmospheric covering. It is interesting to see how completely these expectations about planetary atmospheres, based on knowledge of surface gravity and of kinetic velocity, are fulfilled by telescopic and spectroscopic observations.

Since those first dramatic moments in 1610 when Galileo turned his telescope upon the planets and learned something of their true nature, there has been hardly a night when one or more observers have not carefully scrutinized the disks of the principal planets with a great variety of optical equipment, but with the eye as the final detector and interpreter. In spite of all the recent advances in astronomy, it is still true that a good eye at the end of a telescope of moderate size can detect finer detail than any photographic process.

This is not the place to give a complete description of all the surface features which have been recorded on the planets. It will be sufficient to mention only the most important.

Mercury undoubtedly shows permanent dark markings of some kind, but the best observers disagree as to their exact form. Venus not infrequently shows faint patches darker than the brilliant background of the disk. These are temporary and change from day to day.

Mars is the only planet whose surface can be studied in detail. The polar caps which melt in spring and reform in autumn are striking features. The reddish background of the planet is interrupted by large blue-green areas which darken as the Martian summer advances. A network of dark linear markings (the "canals" of Schiaparelli) connect the blue-green areas. Different observers give different accounts of the canals, some reporting them as narrow and extremely straight while others find them diffuse in outline. Evanescent white clouds have often been observed, proving beyond a doubt that the surface of the planet is covered by a thin gaseous atmosphere.

Much detail is visible on Jupiter under favorable conditions. The surface is covered by a heavy blanket of clouds among which appear the everchanging reddish-brown belts roughly parallel to the equator. The edge of the planet is much fainter than the center. Saturn presents a similar picture, apart from its unique system of rings, but the belts are not so clearly marked as in the case of Jupiter. The surface of Uranus shows very little. Its color is definitely greenish (owing to absorption of red light by methane in the atmosphere) and some observers have detected faint bands. Neptune is a still less satisfactory object. The green color is more pronounced, but no detail can be detected with certainty.

Photography

The invention of photography has not had quite the same revolutionary effect on the study of planetary surfaces that it has had on stellar and nebular investigations and on spectroscopy.

In the case of stellar work, much can be done without insisting on perfect definition of the image, because blurring does not change measures of either its brightness or its mean position. Any loss in sharpness is largely made up for by the ability of the photographic plate to integrate the light which falls on it over long periods of time. And, except in the case of double stars, clusters, and nebulæ, there is no detail to be lost through poor atmospheric conditions, because the stars are so distant that they would appear only as the diffraction images of points of light, even in the largest telescopes under perfect conditions.

But planetary work is far more exacting. A study of minute detail on the surface of a planet requires that points very close to one another be clearly distinguished, and this means that the light from each point must be sharply concentrated, without serious blurring which would cause partial overlapping of images and loss of recognizable detail.

Although theoretical optical resolution increases in direct proportion to the diameter of the telescope, a practical limit to the attainable resolution is usually set by characteristics of the photographic emulsion and by turbulent conditions in the atmosphere through which the light must pass. Because of atmospheric disturbances the largest telescopes will seldom show finer detail than will an instrument having an aperture of 20 or 30 inches.

Although photography can rarely show more than the eye can see, it has two marked advantages. One is the permanent and unbiased record which it supplies. The other is its ability to portray the surface of a planet in light of different colors. A comparison of exposures made in different regions of the spectrum has resulted in much valuable information as to the nature of planetary atmospheres and as to the various levels at which visible features are located.

By far the most extensive collection of planetary photographs in existence has been made by E. C. Slipher at the Lowell Observatory, using a negative enlarging lens near the focus of the 24-inch refractor and of the 40-inch reflector. Other important work has been done by Barnard at the Yerkes Observatory, by W. H. Wright at the Lick Observatory, and by R. W. Wood and by Ross and others at Mount Wilson. Planetary photographs have recently been made at the coudé focus of the 100-inch telescope where the effective focal length is 250 feet, so that no enlarging lens is needed.

Advantage has been taken of the achromatic image produced by the coudé mirror system to attempt color photographs of Mars, Jupiter, and Saturn with Eastman Kodachrome film. Apart from its rendering of color, this film has the advantage of freedom from photographic grain.

Mercury is a difficult object to photograph because it always appears close to the sun. Few definite markings have been recorded and the difference is not great between photographs made with light of different colors.

Venus is more satisfactory. Ultraviolet photographs made at Mount Wilson by Ross (plate 2, figure 1) show dark irregular markings of a temporary nature. Their failure to show consistent motion suggests that the planet's period of rotation must be several days at least. These markings do not show in light of longer wave length, which indicates that they are irregularities in the highest layers of the atmosphere. Attempts to pierce the layer of cloud or dust which surrounds the planet, and to photograph the surface with infrared light have not yet been successful.

When Mars is photographed with ultraviolet light little is ordinarily seen except the brilliant white polar cap and the faint rim light. Not infrequently thin clouds can be detected. On rare occasions these are prominent features, as was the great white cloud, 400 by 800 miles, detected by E. C. Slipher on July 9, 1922. A sudden clearing in the atmosphere was noted by the same observer in May 1937. For several days his blue photographs showed almost as much detail as red photographs. This marked clearing indicates that the normal opacity to ultraviolet light may be as much as twice that of the earth's atmosphere.

Photographs of Mars made with red light are very different. They show much more fading toward the edge of the planet and the dark blue-green areas stand out in sharp contrast against the bright background of the reddish-ochre areas, some of which nearly equal the polar cap in light intensity. A number of canals have been photographed, but their detailed characteristics can be studied only visually and by experienced observers.

Although Jupiter is heavily covered with clouds, so that the actual surface is quite invisible, there is a wealth of detail in the clouds themselves (plate 1). The tropical belts show almost as well in blue light as they do in yellow light, and in blue photographs they extend to the edge of the planet, suggesting that they lie at a high level in the atmosphere. The great red spot is extremely dark in blue and ultraviolet photographs. In red photographs some components of the belts, which appear red to the eye, disappear entirely, owing to lack of contrast with the background of the ball. The general fading of the planet toward the edge is much more marked in red light than in blue light. But W. H. Wright has called attention to the fact that the polar areas are fainter in the ultraviolet, so that the ellipticity is exaggerated as compared with red photographs in which the planet is more nearly round. The belts show much more detail in red light than in blue.

Saturn is in most respects similar to Jupiter, except that the belts and surface features are all much less prominent. E. C. Slipher noted a marked change in the color of the equatorial region between 1927 and 1929, during which interval the surface became so red that blue photographs were dark in comparison with those made with yellow light. R. W. Wood noted a similar dark equatorial band in 1915. In 1930 the two kinds of photograph were almost equal in density.

Since Uranus and Neptune appear definitely greenish to the eye, the red photographs, as might be expected, are definitely weaker than the blue photographs. Both planets are so distant that little detail can be recorded.

The disk of Pluto has never been seen. Baade at Mount Wilson has measured the brightness of the planet with the 60-inch and 100-inch telescopes and finds that its magnitude must vary between 14.14 and 16.39, according to its position in its highly eccentric orbit. The color index corresponds to a G-type star, slightly more yellow than the sun. The brightness of Pluto is almost the same as that which Triton, the satellite of Neptune, would have if the two were at the same distance, indicating that they are probably much the same in size.

PLANETARY TEMPERATURES

There are two principal problems to be solved in measuring planetary temperatures. The first arises from the fact that the radiation received from any of the planets consists of two parts: reflected radiation from the sun and true planetary radiation characteristic of the temperature of the emitting surface. Fortunately, the maxima of intensity for these two kinds of radiation are so widely separated in the spectrum that they scarcely overlap, one being in the visual region and the other far out in the infrared. The two kinds of radiation may therefore be separated by an appropriate filter.

A thin microscope cover glass absorbs all wave lengths longer than about 6 μ, while a water cell cuts off radiation near 1.3 μ. Since the planetary heat which has passed through the earth's atmosphere is limited almost completely to the region between 8 μ and 14 μ, its amount is readily determined by taking the difference between radiometric measures made with one or the other of these filters and without any filter at all.

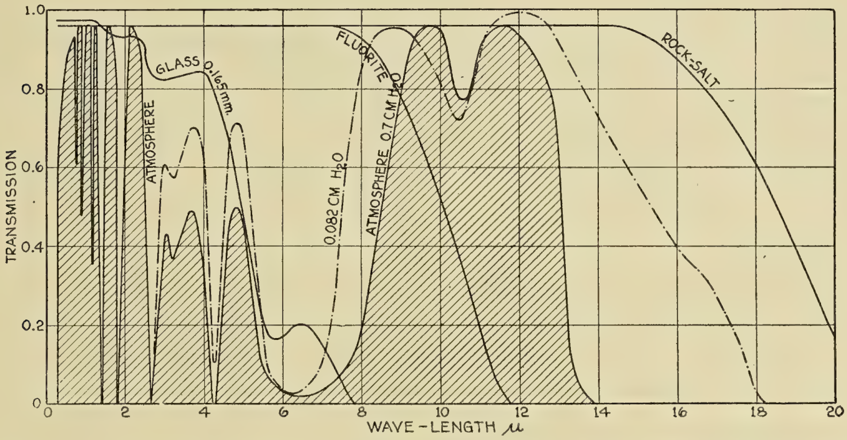


FIG. 1. Transmission curves of (*a*) the atmosphere above Mount Wilson (the shaded curve); (*b*) water vapor, 0.082 cm precipitable water; (*c*) microscope cover-glass, 0.165 mm thick; (*d*) fluorite, 4 mm; and (*e*) rock salt, 2 mm.

The other problem is that of correcting for the absorption in the earth's atmosphere, which is due primarily to water vapor. This correction requires a knowledge of the total amount of water vapor above the observing station, determined from the vapor pressure at the time of observation. It also requires an exact knowledge of the degree of absorption at each wave length caused by a known amount of water vapor (figure 1). The laboratory measures of Fowle, which are the best available and refer to 0.082 cm of precipitable water, must be extrapolated to 0.7 cm, the average amount of precipitable water encountered in a dry climate. By combining theoretical black-body emission curves with this absorption curve, it is possible to compute what fraction of the energy emitted by a planet at any temperature will pass through the earth's atmosphere and reach the radiometric instrument. This last step involves a summation of the products, at a series of wave lengths, of the intensity of black-body radiation multiplied by the atmospheric absorption.

The earliest measures of lunar radiation were made by the Earl of Rosse, using a thermopile in the focus of his large reflecting telescope, from which he concluded that the temperature at the center of the full moon was probably as high as that of boiling water. Further measures, leading to similar results, were made by Langley and by Very with a bolometer before the beginning of the present century.

Vacuum thermocouples were used by Coblentz in 1914 at the focus of the 36-inch refractor of the Lick Observatory. Coblentz and Lampland have made extensive measures of planetary radiation at the Lowell Observatory, with the aid of the 40-inch reflector. Menzel, in 1923, discussed the theoretical aspects of the problem and used the data given by Coblentz and Lampland to derive actual temperatures.

At Mount Wilson, an extensive series of observations on lunar and planetary radiation has been made by Pettit and Nicholson, using a vacuum thermocouple at the focus of the 100-inch reflector. The large aperture of this instrument made possible the use of receiving surfaces small enough to permit the investigation of different regions on the surface of a planet. Planetary radiation has been separated from reflected solar radiation by the use of either a microscope cover glass or a water cell. The measures of radiation have been placed on an absolute basis by direct comparison with stars of known radiometric magnitude. It has thus been possible to express the radiation from each planetary surface in terms of its radiometric magnitude per square second of arc. After allowing for the length of air path and other minor corrections, the temperature was obtained by comparing this value with the radiometric magnitude per unit solid angle computed for surfaces at a variety of temperatures after their radiation had passed through the atmosphere.

The temperatures of different points on the surface of the moon were measured by Pettit and Nicholson at the focus of a 27-inch mirror of 67 inches focal length, fed by the cœlostat of the Snow telescope. They found that the temperature of the central point on the full moon, where the sun is in the zenith, is 101°C., while the temperature of points near the edge of the full moon, where the sun is close to the horizon, is about −50°C. When the subsolar region is seen edgewise, as it is by observers on the earth at first or third quarter phase, its apparent temperature is only about 50°C. This fact is probably the result of the rough surface, which permits more radiation to escape toward the zenith than toward the horizon. One other result is of particular interest: the striking quickness with which the temperature of the surface of the moon was observed to cool from close to that of boiling water down to about −120°C., in a period of less than three hours, when the sun's heat was suddenly cut off at a total eclipse of the moon on June 14, 1927. This rapid cooling indicates a very low conductivity of the outermost layers of our satellite, and suggests substances such as pumice or volcanic ash. Measures of polarized light reflected from the moon, made at Mount Wilson by F. E. Wright, lead to the same conclusion.

A determination of the temperature of Mercury was made by Pettit and Nicholson, using the 100-inch telescope in full daylight. The high temperature of the planet leads to the emission of so much radiation on the short wave-length side of 6μ, where the cover glass cuts off, that a water cell (whose limit is 1.3μ) was in this case a more satisfactory filter for separating planetary heat from reflected light. The image of the planet was too small to permit measurements on individual parts of the disk. A plot of the radiometric magnitude of planetary radiation and of reflected light (reduced to mean distance from the sun and unit distance from the earth) against phase angle was very similar to such a plot for the moon. The temperature of the subsolar point was found to be about 327°C. when the planet was at mean distance from the sun.

Measures of radiation from the dark surface of Venus at crescent phase have been made by Pettit and Nicholson, and show that the temperature is about −23°C. The temperature of the illuminated portion of the disk seemed not to be very different. Coblentz and Lampland, however, derived a temperature of 60°C. for the illuminated surface but agree in finding no marked difference between this and the dark portion. These results, although not yet final, are highly significant, for they show almost conclusively that the planet cannot keep the same face always toward the sun, which would imply a period of rotation of 225 days. On the other hand the spectroscopic observations show that the period cannot be less than about 6 days. The observed temperatures are entirely consistent with radiation reflected from a layer of high cirrus clouds on a slowly revolving planet.

The temperature for the whole disk of Mars is found to be about −23°C. At noon in the equatorial region the temperature rises a little above freezing. The polar caps have a temperature of about −70°C. and the limb about −13°C. During the night the temperature must fall to a very low value. Coblentz and Lampland made extensive measures with almost the same results. They find that at the center of the disk the temperature of the bright yellowish-red desert areas is about −5°C., while that of the dark greenish areas in some instances is as much as 22°C. Pettit and Nicholson's drift curves, taken across the disk from east to west, are symmetrical and follow the sine law, as they do for the moon and Mercury, with little if any displacement of the maximum from the noon position. This indicates a rare atmosphere. The observed temperatures are more suggestive of snow than of solid carbon dioxide as the material composing the polar caps.

Great difficulties are encountered in attempting to measure the temperatures of the visible layers on the major planets. The temperature is so low that the thermoelectric current to be measured is very small, but even more serious is the fact that the portion of the total radiation transmitted by our atmosphere between 8μ and 14μ is a small fraction of the total planetary radiation. The maximum of intensity is at about 20μ, where the atmosphere is utterly opaque. Absorption by methane in the planetary atmospheres is another complicating factor. Pettit and Nicholson have found a temperature of about −138°C. for the surface of Jupiter. From the data obtained by Coblentz and Lampland, Menzel has computed temperatures as follows: Jupiter −135°C., Saturn −150°C., Uranus less than −185°C. The temperature of Jupiter agrees closely with the minimum set by spectroscopic determinations of the amount of gaseous ammonia. It is significant that the measured temperatures of the outer planets agree reasonably well with what would be expected from solar radiation falling upon them. There is no evidence that internal heat makes any important contribution to the total radiation.

PLANETARY SPECTRA

Two kinds of planetary investigation have been made possible by the spectrograph. One is the question of the period of rotation of those planets which do not show features definite enough to permit a visual determination of the rate of rotation. No satisfactory observations have yet been made on Mercury, and its period is still dependent on conflicting visual observations and on radiometric evidence.

An attempt to derive a spectroscopic period of rotation for Venus was made by Slipher in 1902 with negative results, and this was later confirmed on plates taken by St. John and Nicholson at Mount Wilson with the 18-foot spectrograph of the Snow telescope.

The periods of Mars and Jupiter can be determined without difficulty by visual observations of well-defined markings, but in the case of Saturn this can be done only occasionally when a spot with sufficient brightness appears in the cloud layer which surrounds the planet. The spectroscopic method was first applied to Saturn by Keeler at Allegheny in 1895; he found the spectrum lines inclined when the slit of his instrument was set along the planet's equator. His period of 10¼ hours agrees well with such visual observations as have been made, and has been independently confirmed by others. Keeler also proved that Saturn's rings are in fact a flattened swarm of satellites, for the spectrograph showed that the inner edge of the ring revolved more rapidly than the outer edge, and that the actual velocities corresponded exactly to those of satellites at the same relative distances.

A period of 10.83 hours for the rotation of Uranus was established by Lowell and V. M. Slipher in 1912, which was strikingly confirmed by Moore and Menzel in 1930. In spite of the retrograde motion of the satellite of Neptune, Moore and Menzel have found that Neptune itself rotates in the direct sense with a period of about 16 hours.

The other field in which the spectrograph has made important additions to our knowledge of the planets is that of the chemical analysis of their atmospheres. Two difficulties are encountered at the outset, however. In the first place, none of the gases likely to exist free in the atmosphere of any planet has its most important spectrum lines in the region available to photography. The well-known oxygen bands in the near infrared are "forbidden" transitions, while the bands of most other gases likely to be encountered are high vibrational harmonics with extremely small absorption coefficients. The second difficulty is connected with the earth's atmosphere, through which observations must be made, for our own atmosphere contains oxygen and water vapor in large amounts and the bands of these molecules are necessarily superposed on the spectrum of every astronomical object. This situation would render hopeless the detection of smaller amounts of either of these gases on any planet but for the clever device first suggested and tried by Lowell and Slipher in 1905. When a planet is so situated in its orbit that it is rapidly approaching or receding from the earth, every line in the bands of oxygen or water vapor which may exist in its atmosphere will be displaced toward the violet or toward the red, and will be at least partially separated from the corresponding telluric line which at other times covers such a planetary line completely. If the planetary line is not actually resolved, it should at least displace the wave length of the telluric line. The same method was independently invented by Campbell and Albrecht at the Lick Observatory in 1910, and has been used extensively at Mount Wilson.

Since Mercury always appears close to the sun, its spectrum can be photographed only in the daytime. The brilliant sky background makes it impossible to do this in any region but the infrared, where the sky is relatively dark. An exploration of the spectrum made by Adams and Dunham has failed to show any planetary bands.

Lowell and Slipher photographed the spectrum of Venus with a 3-prism spectrograph (50 A per mm) in 1902 and compared the wave lengths of the water-vapor lines in the yellow, and of the lines in the red α band of oxygen, with the wave lengths of the corresponding lines in spectra of the moon. The results were inconclusive as to the existence of either gas.

St. John and Nicholson, in 1921, used much higher dispersion (3 A per mm) on these same lines, and also on lines in the B band of oxygen. The spectra were obtained with the 18-foot Littrow spectrograph of the Snow telescope on Mount Wilson and a bright first-order grating ruled by Anderson at Johns Hopkins. No resolved components due to the atmosphere of Venus could be detected and there were no shifts in the wave lengths of the oxygen and water-vapor lines. These observers concluded that

> in the path traversed by solar light through Venus' atmosphere there must be less than the equivalent of one meter of oxygen, less than one-thousandth of that in our atmosphere, and less than one millimeter of precipitable water vapor.

Two new bands in the infrared spectrum of Venus were discovered in 1932 by Adams and Dunham on photographs taken with the 9-foot spectrograph at the coudé focus of the 100-inch reflector. These bands (plate 2, figure 2) had never been observed in any laboratory source, and their identification was made on the basis of the moment of inertia of the molecule, determined from the spacing of individual rotational lines. Both bands gave a value agreeing closely with that of carbon dioxide. This identification was verified by finding a third band whose existence was predicted at a still longer wave length in the infrared. The bands are vibration-rotation combinations between the ground level and the $5\nu_3$ and the $5\nu_3 + (\nu_1, 2\nu_2)$ levels.

As soon as the presence of carbon dioxide had been established on Venus, it became of considerable interest to estimate the amount of the gas from the intensity of the observed band. In order to do this, as well as to establish the identification beyond doubt, a closed pipe 20 m long was set up by the writer in the building of the Snow telescope. Light was sent through this pipe twice by reflection, giving a total path of 40 m, after which it was focused on the slit of a spectrograph comprising the same optical parts as were used in photographing the spectrum of Venus. No absorption could be detected when the pipe was filled with carbon dioxide at atmospheric pressure, but when the pressure was increased to 10 atmospheres a very faint broad line appeared at exactly the same position as the head of the short wave-length band in Venus. A comparison of the strength of the absorption in the two sources indicated that the amount of CO_2 in the path traversed by sunlight in the atmosphere of Venus must be about 400 meter-atmospheres. Pressure effects, which seriously vitiate this result, can be eliminated only by making the laboratory comparison at lower pressure and in a far longer pipe. Adel and Slipher from an examination of published photographs have estimated that the amount of carbon dioxide is two mile-atmospheres, and Adel has attempted to estimate the temperature of the gas from the relative strength of lines in the bands. Accurate determinations of these values depend on reliable photometric measures of the Venus bands, for which the necessary observations have recently been obtained at Mount Wilson.

The first attempt to detect oxygen and water vapor in the atmosphere of Mars was made by Lowell and Slipher in 1902 at the time of their study of Venus. Their conclusion, and that of Campbell and Albrecht who used the same method, was negative. A comparison of the integrated intensity of the B band of oxygen in the spectra of Mars and of the moon at the same altitude was made in 1909 by Very, who concluded that the Martian atmosphere contains about one-fourth as much oxygen as does our own atmosphere.

The first measures at Mount Wilson were made on spectra photographed by Adams and St. John in 1926. These spectra had a linear scale of 11 A per mm, produced by the use of six prisms, and when tracings of the oxygen lines in the α band were studied with a microphotometer there was evidence of a slight displacement, corresponding to about 16 per cent as much oxygen on Mars as on the earth.

In 1932, a new photographic emulsion was made available by Dr. Mees of the Eastman Kodak Company with high sensitivity extending beyond λ6800. This emulsion enabled Adams and Dunham to photograph the region of the B band of oxygen with the 9-foot grating spectrograph at the 100-inch telescope. Since fewer molecules are required to produce detectable absorption of the B band than of the α band, this provided a more delicate test for oxygen than had been possible with earlier photographic emulsions. Spectra of Mars (plate 3, figure 1) were obtained during its approach to the earth in November 1932 (velocity = -13.75 km/sec) and its recession in May 1933 ($+12.42$ km/sec). Although the velocity of the planet was large enough to separate almost completely any lines due to Martian oxygen from the telluric lines, no such lines could be detected. An investigation of the contours of individual telluric oxygen lines failed to reveal an asymmetry as great in amount as would be caused by Martian oxygen if it existed to the extent of 1 per cent of the quantity in the earth's atmosphere. An investigation of the water-vapor band at λ7200 was also negative.

In spite of this spectroscopic evidence against the presence of oxygen, it is more than likely, as Russell has pointed out, that it has existed in the past, and that later it combined with ferrous iron to form ferric compounds with a reddish color similar to that of the desert areas on the planet at the present time. It is almost certain, moreover, that the white polar caps are composed of ice or snow, since the temperature is too high to permit carbon dioxide to remain very long in the solid state. It must be admitted that the visual observation of a thin sheet of ice covering a limited area of the planet is a more delicate test for water than is the most powerful spectrograph of the present day, because the spectrograph can detect only the vapor and this may be distributed through a large part of the entire atmosphere. It would therefore be unwise to say that there may not be enough of both water vapor and oxygen to support life in some form which may have become gradually adapted to the rigorous conditions existing on Mars.

The spectra of the four major planets, Jupiter, Saturn, Uranus, and Neptune, present an entirely different problem. Here it is much less a question of searching for bands of vanishing intensity than it is a question of identifying bands of such tremendous absorbing power for red light that they completely alter the color of the outer planets and cause the characteristic greenish tinge of Uranus and Neptune.

The bands in the spectrum of Uranus are so conspicuous that they were detected visually by Secchi and Huggins many years ago. An admirable series of photographs of spectra of the major planets was published by V. M. Slipher in 1909, reaching to the end of the visible red, and this wave-length limit has been greatly extended with the development of new photographic emulsions in recent years. The spectra of all four of the major planets show a series of wide bands, beginning in the green and growing stronger in the red, and stronger yet in the infrared. Most of the absorption bands are somewhat heavier in Saturn than in Jupiter, and in Uranus and Neptune they are so strong as nearly to obliterate the infrared spectrum.

The origin of these bands was unknown for many years. In 1932, Wildt at Göttingen pointed out the agreement in position of the band at λ6450 with an ammonia band measured by Badger, and of the bands at λλ5430, 6190, 7260, and 8800 with known bands of methane.

Improved photographic emulsions had at the same time enabled the writer to obtain spectra of Jupiter and Saturn with high dispersion, using the 100-inch coudé spectrograph and a very bright grating. These spectra resolved the planetary bands into their individual component lines, so that a direct comparison with the fine structure in the spectra of various laboratory gases became possible. For this purpose the same absorption pipe which had been used to verify the presence of carbon dioxide on Venus was available. The absorption spectrum of 40 m of ammonia at atmospheric pressure showed (plate 3, figure 2) an almost perfect duplication of the Jovian bands at λ6450 and λ7920. In the first of these bands 30 lines due to ammonia could be identified, while in the second band 39 lines were represented. Photographs of the spectrum of ammonia were taken under reduced pressure to match the intensity of the band lines in the planet. As a result of this comparison it appears probable that the amount of ammonia above the reflecting layer on Jupter is equivalent to about 10 m of the gas at atmospheric pressure. The fine structure of the planetary bands shows differences in relative intensities as compared with laboratory bands. These differences are probably caused by the much lower temperature on Jupiter, and it is hoped that measures of the individual lines may provide data for an independent determination of the temperature. Both of the ammonia bands are much weaker in the spectrum of Saturn than they are in that of Jupiter.

A minor limit for the temperature at the cloud layer on Jupiter may be inferred with some accuracy from the amount of ammonia gas which is observed above this level. The partial pressure of this amount of ammonia on a planet of known surface gravity can easily be computed. This pressure equals the laboratory value for solid ammonia at −103°C. The temperature at the cloud level can hardly be less than this or some of the ammonia gas would condense and more of the solid phase would form.

It is more than likely that the atmospheres of the outer planets contain great quantities of hydrogen, derived from the sun and retained by their strong gravitational action. The mixing of an excess of hydrogen with ammonia would reduce the partial pressure of the latter to such an extent that the temperature could be as low as −123°C. and yet the observed amount of ammonia would be supported.

The agreement of the radiometric temperature of Jupiter with these temperatures for equilibrium between the observed amount of ammonia and the solid phase make it quite reasonable to suppose that the clouds actually contain crystals of ammonia. The relative weakness of the ammonia bands in the spectrum of Saturn finds a ready explanation in the lower temperature, which would cause more condensation of the gas.

High-dispersion spectra of Jupiter and Saturn obtained with the 100-inch spectrograph also show excellent agreement with the spectrum of methane photographed in the absorption pipe on Mount Wilson. Eighteen of the strongest lines in the band near $\lambda 8640$ can be identified with certainty. Slipher and Adel have recently shown that methane accounts quite satisfactorily for all the other bands so far observed, and they estimate that the amount of this gas ranges from 0.5 mile-atmosphere on Jupiter to 25 mile-atmospheres on Neptune.

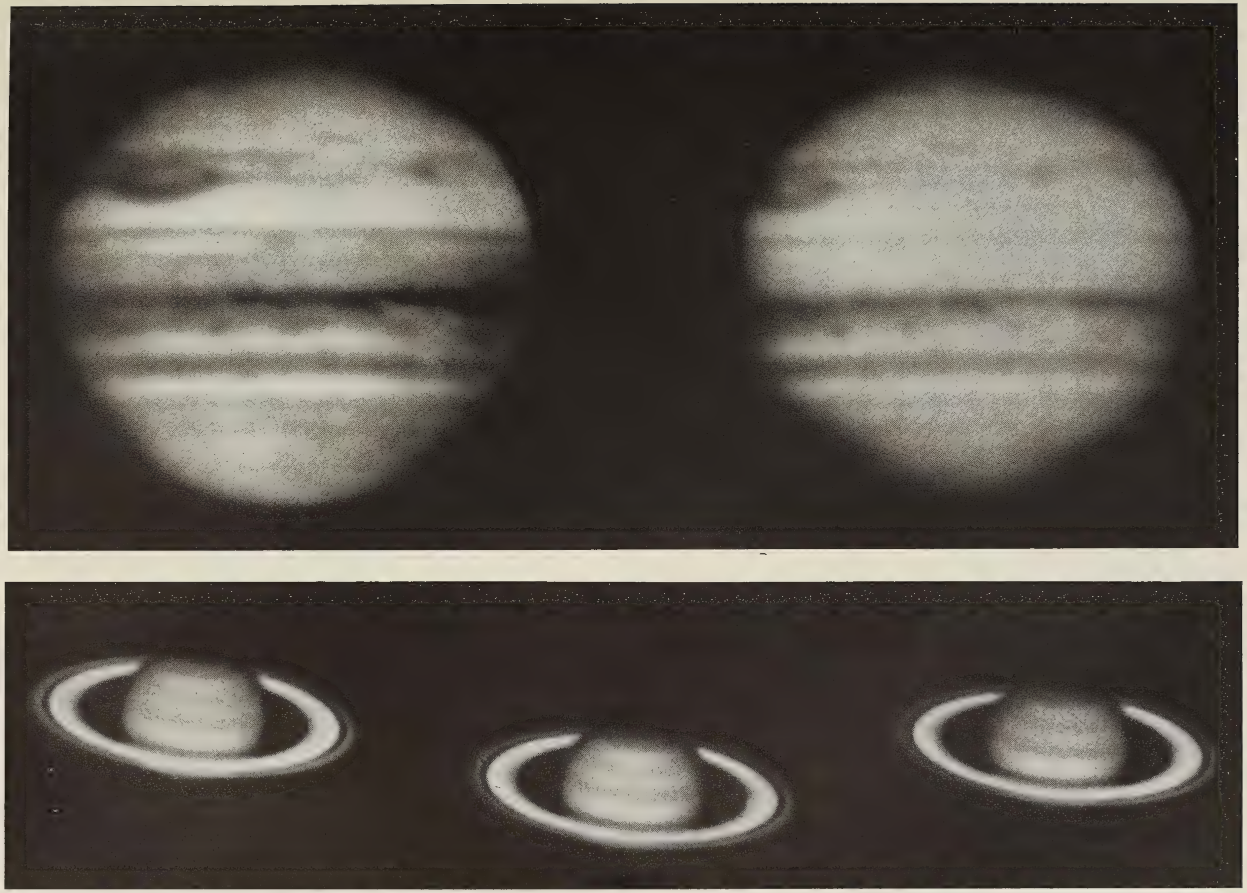
The composition of the atmospheres on all the planets, so far as it is now known, is exactly what might have been predicted in advance. The major planets with their high force of gravity retained hydrogen as it came to them from the sun. Much of it may still be there as an excess of free gas in their atmospheres. The rest combined with the greater part of the oxygen, nitrogen, and carbon as the planets cooled. The simplest compounds which could form are water (H_2O), ammonia (NH_3), and methane (CH_4). The water probably froze to ice and fell beneath the clouds. The ammonia, in frozen form, may well constitute a considerable portion of the clouds. Enough to be detected lies, in the gaseous form, above the clouds on Jupiter, less on Saturn where it is colder, and still less on Uranus and Neptune. Russell suggests that the freezing out of ammonia on the most distant planets explains the greater depth of visible methane in their upper atmospheres.

The earth and Venus, and probably Mars (judged from the polar caps and the red color), have atmospheres containing oxygen or compounds of oxygen. This has been made possible only by the past escape of hydrogen, which, if present in excess, would have tied up the oxygen as water. The absence of any atmosphere on Mercury, the satellites, and the minor planets has been explained on thermal and gravitational grounds.

CONCLUSION

Much remains to be learned about the planets. This brief survey merely shows with what certainty new knowledge follows the invention of new equipment and new methods for analyzing and interpreting the radiation which comes to us from these members of the solar system.

What is the period of rotation of Mercury and Venus? Is there a trace of oxygen and water vapor on Venus and Mars? What is the principal constituent of the atmosphere of Mars? Is the surface of Venus obscured by ordinary clouds or by dust, and can this layer be penetrated by long-wave radiation so as to reveal the nature of the actual surface of the planet? What is the exact law of fading toward the limb in light of various wave lengths for each planet, and what can it reveal about the physics of the atmospheres? What chemical compounds are responsible for the brilliant colors in the bands of Jupiter? These are merely a few of the questions still waiting to be answered.



Upper: Two views of Jupiter taken at Mount Wilson Observatory, the one in ultraviolet light, the other in blue-violet light. It is almost certain that the belts represent clouds, probably not of water but of some substance that vaporizes at a lower temperature.

Lower: Views of Saturn taken at Mount Wilson Observatory with the 60-inch telescope. The concentric rings are composed of swarms of little satellites, each swarm pursuing its own independent orbit around the planet.

Plate 2

Fig. 1. Photographs of Venus taken by Dr. F. E. Ross in 1927 with the 60-inch and 100-inch reflectors on Mount Wilson, using ultraviolet light. The upper series was taken in June, the lower in July. The dark markings are not permanent but change from day to day.

Fig. 2. Spectrum of Venus in the infrared, showing absorption bands due to carbon dioxide, oxygen, and water vapor. Venus was approaching the earth with a velocity of 13 km a second when these photographs were taken, with the result that any band lines originating in her atmosphere are displaced toward the violet with respect to similar lines originating in the earth's atmosphere. The positions are indicated at which Venus lines and telluric lines would be expected to show. Solid lines have been drawn where the corresponding absorption lines are actually seen; broken lines where no such lines are seen. There is evidence for a large amount of carbon dioxide, but no evidence for oxygen or water vapor in the atmosphere of Venus. This spectrum, as well as those of Mars, Jupiter, and Saturn, were photographed with the coudé spectrograph of the 100-inch telescope.

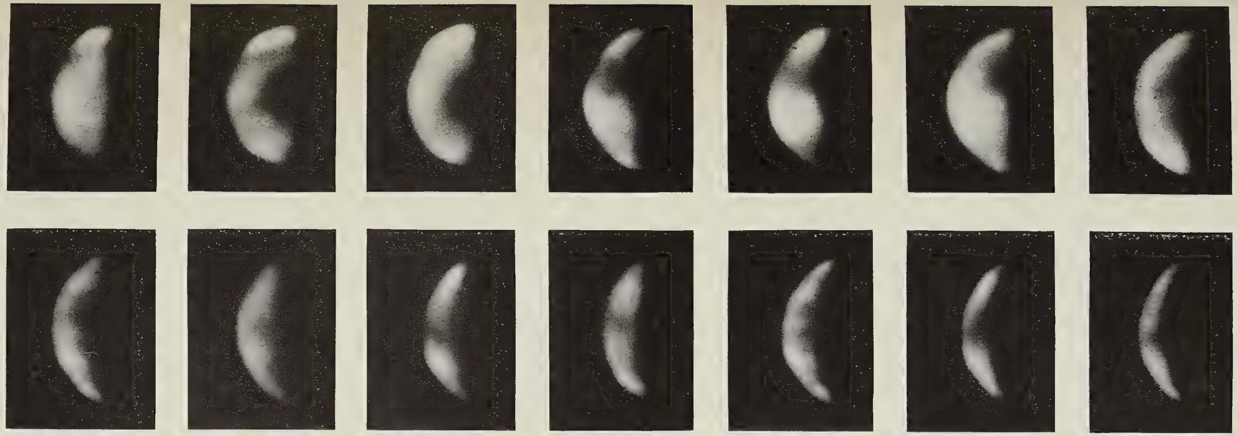
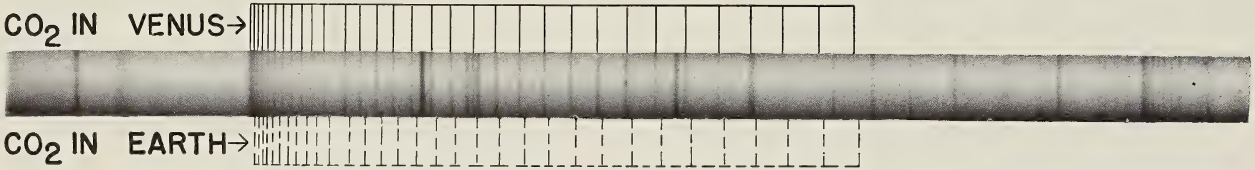
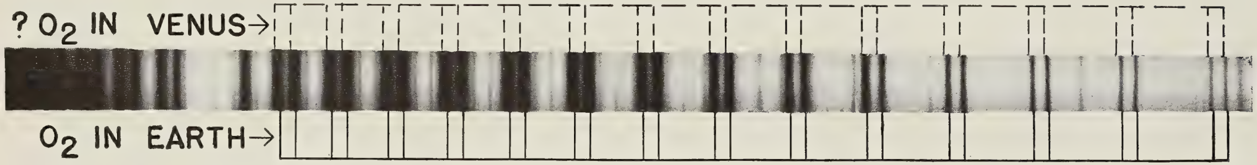


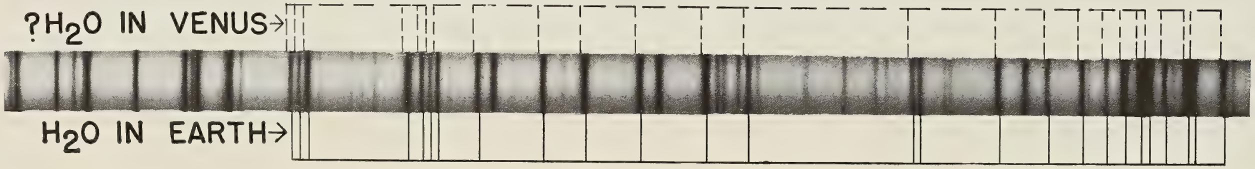
FIG. 1. Photographs of Venus.



Carbon dioxide



Oxygen



Water vapor

FIG. 2. Spectrum of Venus.

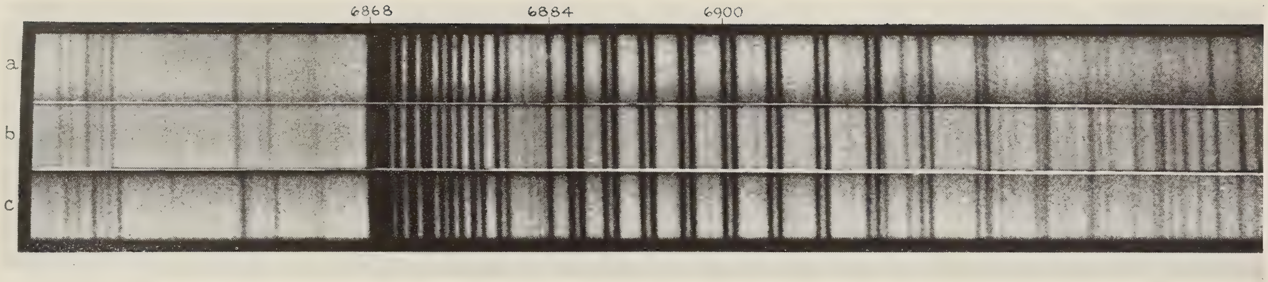


Fig. 1. Spectrum of sun and Mars: (*a*) Mars, November 13, 1932; (*b*) sun; (*c*) Mars, May 3, 1933.

The B band of oxygen in the spectrum of Mars approaching the earth with a velocity of 13.75 km per second (upper spectrum) and receding with a velocity of 12.42 km per second (lower spectrum), compared with the spectrum of the sun (middle spectrum). No components due to Martian oxygen can be found and there is no detectable shift in the measured positions of the lines.

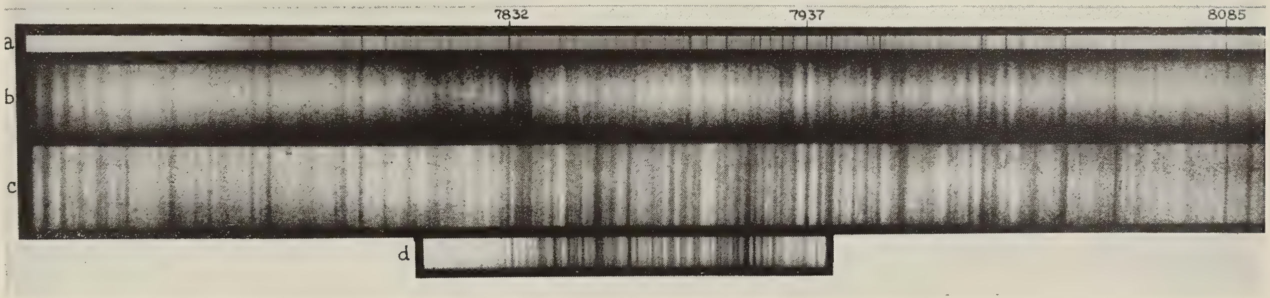


Fig. 2. Near infrared spectrum of Saturn and Jupiter: (*a*) sun; (*b*) Saturn; (*c*) Jupiter; (*d*) ammonia gas.

In comparing the spectra of Saturn and Jupiter with the laboratory spectrum of ammonia it is found that there is less ammonia on Saturn than on Jupiter.

SUNSPOTS AND STELLAR DISTANCES

WALTER S. ADAMS

Director, Mount Wilson Observatory
The Carnegie Institution of Washington

To the investigator in the field of physical science come two very great intellectual satisfactions. The first is the pure joy of discovery, of finding out some fact or sequence of phenomena which enables him to gain some further insight into the nature of the physical world about him. The second is the stimulating and pleasing uncertainty regarding the distance to which a discovery may carry him or the direction a patient following of its consequences may eventually take. A small fossil remnant may make possible the reconstruction of a prehistoric flora or fauna and enable the scientist to draw the widest inferences regarding the migration of life throughout great periods of time, to trace the earth's geography and to study climatic conditions over its surface; or the discovery of a single spectrum line may lead to a conception of matter in a state never observed in any terrestrial laboratory and enable him to picture the operation of the giant forces in a distant star.

The investigation which is outlined briefly in these pages illustrates both these rewards of research. Its immediate consequences were the discovery and an adequate explanation of numerous complex phenomena in the analyzed light of the changing spots which form over the surface of our sun, increase or decrease in size, and pass across the disk as the sun turns slowly on its axis. Its more distant and far-reaching results were the discovery of the existence of magnetism in the sun, an insight into the nature of the radiation of the atom, and the development of a new and powerful method for determining the distances of the stars. The results attained testify in an unusual degree to the value of a combined attack upon a problem by a number of investigators with varying points of view and different types of training; they have also greatly influenced and expanded the outlook upon numerous problems of astrophysical research, not only at the Mount Wilson Observatory but at many observatories and laboratories throughout the world.

An analysis of the light of the general surface of the sun with a spectroscope shows to the eye a brilliant continuous band of color extending from the far violet to the deep red. Crossing this band are thousands of narrow lines which appear dark against the bright background of the continuous band. This band, or continuous spectrum as it is called, originates in the radiation from the hotter and denser gases of the sun which lie beneath its general atmosphere, while the dark lines originate in the somewhat cooler gases of the atmosphere itself, the atoms of these gases absorbing out of the continuous spectrum narrow lines, the positions and intensities of which depend upon the elements composing the gases. Each element has its own distinctive pattern of lines, and it is the presence of great numbers of elements in the sun's atmosphere which produces the enormous complexity of the solar spectrum. About 60 different elements have been identified in the sun through a comparison of the positions of the dark lines in the solar spectrum with those produced in a physical laboratory when the spectrum of one element after another vaporized in an electric arc or spark is examined with a spectroscope. Not only do the solar spectrum lines enable us immediately to identify the elements present in the sun; they also give extremely important information regarding the temperature and density of the gases, the relative abundance of the different elements, and the turbulent motions taking place throughout the disturbed regions of the sun's atmosphere.

When the image of the sun formed by a telescope is moved so that the center of a sunspot falls upon the slit of the spectroscope some conspicuous differences from the solar spectrum are seen at once. In the first place, the band of continuous spectrum becomes much fainter, the effect increasing progressively from the red toward the violet end of the spectrum. In the second place, notable changes occur in the intensities and character of many of the spectral lines when compared with those on the sun's disk, some being greatly strengthened, some weakened, and a few characterized by strong wide wings on either side of the lines. Finally, under the best conditions of observation with a spectrograph giving excellent resolution, some of the lines are seen to be double and great numbers of faint lines, arranged in the form of bands, appear in the spectrum of the spot. Most of these facts had become known through the visual observations of C. A. Young, Lockyer, Fowler, and others; but a great advance was made when it became possible to apply photography to the study of the sunspot spectrum and to utilize fixed spectrographs of great size in combination with a large image of the sun. Dr. Hale was a leader in this development, and with the establishment of the Mount Wilson Observatory, he found a favorable opportunity to undertake an extensive investigation under excellent observing conditions.

The first observations of the spectrum of sunspots were made at Mount Wilson in 1906 by Hale and Adams. A study of the photographs provided a wealth of material giving the identification of many hundreds of lines and the extent to which lines of the various elements are strengthened or weakened in spots when compared with the normal solar spectrum. Accurate measurements of the faint lines appearing in bands led to the identification of the spectra of two compounds, titanium oxide and calcium hydride; and a third, magnesium hydride, was added through the investigations of Fowler in England. In these cases the observed bands are due to the absorption of the molecule of the compound and are termed molecular spectra as contrasted with the spectra of individual lines, which are due to the absorption of the atom. The presence of bands is an unfailing indication of relatively low temperature, since molecules of compounds are dissociated into atoms at temperatures like that of the general surface of the sun or higher.

The most significant result, however, as bearing on the future development of this investigation was the striking fact that among the lines due to the atom of a single element, such for example as iron, many lines were greatly strengthened in the spectrum of the spot, while many others were but slightly affected, if at all. It was clear that some cause was operating to differentiate the behavior of lines of the same element, and the discovery of this cause became the next important step in the investigation.

The simplest and most natural working hypothesis was that the observed effect was due to the lower temperature of the spot as compared with the solar surface. The relative darkness of spots and the presence of molecular spectra made it highly probable that their temperatures were low, but it was by no means certain that a difference in temperature would be adequate to explain the complicated phenomena observed in the spectrum. At this period in the development of spectroscopy little was known regarding the effect of change of temperature upon the intensities of spectral lines. The work of Lockyer and others had shown the great astrophysical importance of "enhanced" lines—lines which are much stronger in the spectrum of the electric spark than in that of the electric arc; but at temperatures below that of the normal arc comparatively little work had been done. The spectrum of the flame of some readily volatilized elements had been studied and it was known that the blue line of calcium appeared at lower temperatures than the great pair of lines in the violet, but our knowledge of the more refractory elements was very slight.

It was clear that results obtained in the laboratory where physical conditions could be defined and controlled must be used for comparison with the sunspot material. Accordingly, an investigation was commenced in the small physical laboratory on Mount Wilson with the joint cooperation of Hale, Adams, and Gale. The spectrum of iron was selected for study at the outset, partly because of the great number of lines it contains, but chiefly because of the remarkable range in behavior shown by its lines, especially in the yellow-green portion of the spot spectrum. The first successful observations consisted of photographs, made on the same plate, of the spectrum of an electric arc with iron poles in which the amount of current was varied through as wide limits as possible. It could fairly be assumed that the greater the amount of current the higher the temperature, although no accurate numerical relationship probably existed because of the variety of other factors. Later, the simpler plan was adopted of comparing the spectrum of the outer presumably cooler portions of the flame of an arc with that of the inner hotter portions near the poles. The results were conclusive in showing a great variety in the behavior of different lines of the iron spectrum, some being relatively much more intense in the outer flame and others in the inner core. Comparison with sunspot spectra showed that just the lines which were strengthened in passing from the spectrum of the sun's disk to that of the spot were those which were strengthened in passing from the core to the flame of the arc. These observations were later fully confirmed with the aid of a simple electric furnace. Similar studies were made of other elements such as titanium, vanadium, and chromium, which have lines conspicuous in sunspot spectra, and both qualitative and quantitative agreement was found between the laboratory and spot results.

Through an unusual coincidence, Fowler at South Kensington carried on in his laboratory investigations very similar to those at Mount Wilson at almost precisely the same time; and the results obtained at the two institutions explained, with but a single important exception, all the prominent characteristics of the spot spectrum. The darkness of spots and the relative weakening of the ultraviolet spectrum when compared with the red, the presence of bands in the spectrum, and, most important of all, the relative behavior of different lines of the same element had been fully accounted for on the basis of a reduction of temperature in spots. The exception was the double character of many lines in the spot spectrum. For some years these double lines had been observed visually, especially by W. M. Mitchell, and with the aid of the powerful spectrographs on Mount Wilson they had been photographed without difficulty. The generally accepted explanation of such lines was that they were due to reversal, a narrow emission line at the center dividing the absorption line into two components. A close study of such lines, however, failed to convince Hale of the adequacy of this hypothesis, and in 1908, influenced greatly by his discovery of the vortical character of sunspots, he undertook tests of the polarization of the components and proved conclusively the existence of the Zeeman effect and the presence of a magnetic field in spots. This brilliant discovery completed the solution of the problem of the sunspot spectrum.

Equally important results, however, developed out of the laboratory investigations themselves. A study of the spectra of different elements photographed in the cooler and hotter portions of the arc made it possible to divide the lines into classes according to behavior with temperature and thus to provide the beginnings of an observational basis for an investigation into the mechanism of the atom producing the lines. King, in the physical laboratory in Pasadena, continued the research with a large electric furnace, in which the temperature could be measured and controlled accurately; and the temperature classes into which he divided the lines of many elements have provided the single most effective means for commencing an attack upon the problem of analyzing the complicated spectrum of any element according to the energy levels of the atom giving out the lines. Modern theory pictures a vast number of configurations in the structure of a radiating atom, and lines produced from the same or closely similar configurations are classed together and behave alike under varying conditions of temperature. Hence the clue afforded by the division of lines into temperature classes through laboratory observations is of immense importance to the unraveling of a complex spectrum, since it shows at once which lines belong to similar states of the atom.

The investigation of the spectrum of sunspots also led to a laboratory study of another class of lines. Observations had shown that "enhanced" or "spark" lines, which, as already stated, are lines which appear only in the spectrum of the spark, or are greatly strengthened in passing from the arc to the spark, are in general weakened when we go from the spectrum of the sun's disk to that of a spot. This might at first thought be regarded as due to the lower temperature of the spot, but this hypothesis is contradicted by the observational fact that in the upper portions of the sun's atmosphere the enhanced lines are relatively stronger than near the general surface of the sun where the temperature must certainly be higher. Accordingly, some cause other than temperature must be sought to explain the behavior of such lines. Experiments were undertaken in the physical laboratory with an electric arc enclosed in a chamber in which different gases could be used, or the pressure and density of any one gas could be varied over wide limits. The resulting photographs of the spectrum showed clearly that the intensity of the "enhanced" lines relative to the unaffected lines increased as the density of the gas surrounding the arc, and hence of the radiating vapor of the element in the arc, was reduced. The very low density of the gases in the upper atmosphere of the sun is, therefore, the reason why the enhanced lines are relatively stronger in this region than near the sun's surface, and the higher density of the gases in spots (well established on other grounds), why they are weaker in the spot spectrum.

These results and many others found an adequate and rational explanation in 1920 when Saha published his theory of ionization. As Bohr and others had shown in previous years, the atom of an element must be regarded as a complex structure consisting of a nucleus and of electrons varying in number according to the atomic number of the element in question. These electrons in a normal atom may be pictured as having a definite configuration. When an element is vaporized in a flame or electric arc or spark, the atoms are excited, electrons move from one configuration to another, and energy is given out in the form of a spectral line. If the exciting source is at a low temperature, the electrons which require little energy to move them, or are in the ground state, as it is termed, are the first to respond, and a series of low-temperature lines is produced. The application of higher temperatures excites electrons in configurations which require more energy to change them and additional lines appear. Finally, if a sufficiently high temperature or other form of excitation is applied, an electron can be completely detached from the original atom and the atom becomes "ionized." The spectrum given out by the ionized atoms of an element usually differs in several marked respects from that given by normal or neutral atoms, the most notable difference being the appearance, or greatly increased intensity, of the enhanced lines. The occurrence of enhanced lines in any spectrum, accordingly, indicates the presence of large numbers of ionized atoms.

One of the chief contributions made by Saha was in showing that low density of a gas favors ionization very greatly. This results chiefly from the fact that when an electron is detached from an atom it immediately seeks to recombine with another ionized atom, to form once more one of the neutral type. If the density of the gas is low, the number of atoms is less and the chances of recombination are more limited so that ionized atoms will remain ionized during a longer period. In the upper atmosphere of the sun, where the density is extremely low, atoms remain in the ionized state for relatively long intervals, and hence many such atoms are available to produce the enhanced lines.

Years before the theory of ionization had been developed, many applications of the results of the early work on the sunspot spectrum and related laboratory spectra had been made at Mount Wilson. In the absence of adequate theory and a rational basis of spectroscopy, some of the conclusions were necessarily based wholly upon the facts of observation, but they were none the less valuable on that account. To mention an illustration from the work of another institution, the whole of the Harvard system of classification of stellar spectra was developed upon an empirical basis and was used by astronomers for many years before its theoretical foundations could be established. Some of the applications of the sunspot results, moreover, were clear from the outset. The presence in a stellar spectrum of the lines and bands prominent in sunspots indicated a temperature for the atmosphere of the star comparable with that in spots, and a relationship between intensity of lines and the temperatures of many stars could be formulated without difficulty. Similarly, conclusions could be drawn regarding the densities of stellar atmospheres as compared with the sun's atmosphere from a study of the behavior and intensities of the enhanced lines. Gradually, as the stellar material increased, these studies began to converge upon problems relating spectral characteristics to the luminosities of stars.

If we consider two stars in the sky of the same apparent brightness, one may be intrinsically very luminous but at a great distance, while the other may be intrinsically faint but relatively near. In other words, the apparent brightness of an individual star may bear no relationship to the total amount of light it gives out, although on the average the apparently brighter stars are also the more luminous. When the distance of a star is known, it is a simple matter to calculate its luminosity from its apparent brightness; and soon after the first successful measurements of distance were made it was found that some stars are vastly more luminous than our sun while others give out much less light.

Now the temperature of a star may be determined with very considerable accuracy from its spectrum, especially from the behavior of the lines and bands which the study of sunspots had shown to be especially sensitive to temperature conditions. If the temperature of two stars is the same, it is clear that the brightness of their surfaces must also be nearly the same. When as the result of observation it was found that two stars of the same spectral type, and hence of nearly the same temperature, differed by a factor of 10,000 in their total luminosities, it became a matter of great interest to make a close comparison of their spectra for possible differences. It seemed hardly possible that such an enormous difference in luminosity should not manifest itself in some way in the quality of the analyzed light of such stars. A study of this question was undertaken at Mount Wilson in 1913 by Adams and

Kohlschütter with the aid of the rich and adequate material on stellar spectra which had been collected with the 60-inch telescope.

At this time the possibility of the loss of light from stars in its passage through space, owing to the scattering and absorption of interstellar dust and gas, was under close examination, especially by Kapteyn and his collaborators. Such an effect would be expected to manifest itself through a loss of blue and violet light relative to red and yellow light, since the scattering by very small particles or by molecules of gas is much greater for light of short than of long wave length. The phenomenon is familiar to us in the blue light of the sky, where the molecules of the gases in our atmosphere scatter the blue light of the sun, and in the red color of the sun when seen through smoke or fog or through the long path in the earth's atmosphere near sunrise or sunset. If then two stars of the same temperature or spectral type are considered, one of which is relatively near and the other very distant, it might be expected that the light of the more distant star would lose more of its blue and violet constituents in the course of its long passage through space and its encounter with scattering clouds of cosmic dust and gas than would the light of the nearer star. The complete investigation of the spectra, accordingly, involved a study of the line spectrum for differences depending upon the luminosities of the stars and a study of the intensities of the blue and violet portions of the continuous spectrum relative to those of the red and yellow portions as dependent upon both luminosity and distance.

Interesting results appeared at once. In the first place, the continuous spectra of stars of high luminosity were found to be notably deficient in violet and blue light when compared with stars of low luminosity of the same temperature. If the intensity of the spectrum in the red region was equalized on the photographs of the spectra of two stars of very different luminosity, the blue portion of the spectrum of the more luminous star faded out much more rapidly than that of the fainter star. This effect dependent upon luminosity proved to be much greater than any effect due to distance in the case of the stars investigated; and as later researches have shown, the amount of selective scattering of light in space for each unit of distance is so small that it can become conspicuous only in the case of stars at very great distances, too great usually for adequate spectroscopic investigation, or of stars embedded in or associated with cosmic dust clouds. So, among stars of the same spectral type, those showing a relatively strong continuous spectrum in the blue region could safely be inferred to be stars of low luminosity and those showing a fainter spectrum, stars of high luminosity.

A still more important result was found from a study of the lines in the spectrum. A close examination of the spectra of two stars of the same spectral type but very different luminosity showed that numerous lines which were prominent in the spectrum of the bright star were relatively faint in the other, and that the converse was equally true. In general, the enhanced lines (due to ionized elements) were most conspicuous in the spectra of the highly luminous stars, and the low-temperature lines in those of lower luminosity; but among these lines a few showed the effect much more conspicuously than did the others. Certain lines due to ionized strontium, for example, were especially prominent in the spectra of the more luminous stars with a temperature equal to that of the sun or lower, and certain lines of ionized iron and titanium in bright stars of higher temperature. In stars of low luminosity, on the other hand, these lines were relatively faint and low-temperature lines of calcium, strontium, and titanium became very prominent, especially in the spectra of the cooler stars of the class. A selection of stars of known distance, and so of known luminosity, made it possible to establish quantitative relationships between luminosity and the measured intensities of these variable lines for stars of different temperatures. As soon as such correlations had been established, it was a very simple matter to reverse the process and in the case of a star of unknown distance to determine the intensities of the lines from the spectrum, to find the corresponding luminosity from the appropriate correlation curve, and then by a very brief calculation to derive the distance of the star.

This method of determining the luminosities and distances of stars has been applied to some 10,000 stars at Mount Wilson and other observatories since its discovery in 1913. Its great advantage consists in the fact that its accuracy is independent of the distance of a star, so that it is most useful in the case of the great majority of stars which lie close to the limit of the effective use of the trigonometric method, in which the small angular displacements of stars resulting from the motion of the earth around the sun are measured directly. The differences in spectral characteristics found in the course of this investigation become so marked in the case of the cooler stars that one of the first results of its application was the full confirmation of the discovery that such stars are divided according to luminosity into two classes known as "giants" and "dwarfs." Giants are stars of great luminosity, and dwarfs, stars of very low luminosity; and the important feature of the discovery is that very few if any stars of intermediate luminosity exist. Based originally by Russell and Hertzsprung upon a very limited number of stars with luminosities determined from trigonometric measures of distance, the theory received immediate confirmation from the much more abundant spectroscopic material. The differences in spectrum between giants and dwarfs among the cooler stars are so marked that a star can be assigned to its class by simple inspection. A spectroscopic study also showed that in stars of different spectral types, from the coolest to those approaching the sun in temperature, this same tendency exists for the luminosities to group around two or three definite values within narrow limits; and this remarkable result constitutes one of the most interesting and significant facts in the problem of stellar formation and evolution.

The development of the theory of ionization gives a simple and quite rational explanation of the behavior of the spectral lines in giant and dwarf stars. If two stars have closely the same temperature as shown by their spectral type, they must have closely the same surface brightness. If, with the same surface brightness, they differ greatly in luminosity, the explanation must be that one star is much larger than the other and has a greater surface from which to radiate light. Now it is well known that the masses of stars do not show a very great range, the vast majority being included within a factor of ten or less. But if the masses of two stars are roughly equal and their luminosities, as observation shows, differ by a factor of many thousands, the brighter star must be very much less dense than the fainter star. In other words, giant stars of low temperature should be huge masses of gas of very low density, and dwarf stars, small compact bodies of relatively high density. The theory of ionization predicts that in the spectrum of a star of low density the enhanced lines should be relatively conspicuous, while in that of a star of high density these lines should be faint and the low-temperature lines strong and characterized by wings. These are precisely the results found by observation, and the empirical method of determining stellar luminosities from the spectrum is fully explained. Finally, the brilliant application of the interferometer to the measurement of stellar diameters by Michelson gave direct objective evidence of the existence of the great red stars of low density inferred from both the ionization theory and the spectrum investigations.

If we summarize this investigation we find three main results: one relates to the sun, one to the physical laboratory, and one to the stars. The first has enabled us to explain the phenomena of sunspot spectra and the solar atmosphere on the basis of temperature and density; the second has made possible the separation of spectral lines into classes according to behavior with temperature, and so to lay an observational basis upon which to build much of the theory of ionization and atomic structure and radiation; the third has given us a simple and direct means of deriving the luminosities and distances of stars with all the wealth of interesting conclusions which such data afford. The completion of the research and the full utilization of its results required the cooperation of numerous investigators: the solar physicist to define and clarify the first stages of the problem; the laboratory spectroscopist to devise the method of experimental attack and interpret the behavior of spectral lines under known conditions of excitation; and the stellar investigator to seek to apply to the almost infinitely varied phenomena of the stars the powerful means of research made available to him. The theoretical physicist then took the results of observation and rounded out the investigation through an adequate and satisfactory theory which has stood the tests of many years.

Although the immediate products of this investigation date back to the earlier years of the Mount Wilson Observatory, their applications have had a permanent and increasing influence upon the research in progress. Spectroscopic determinations of stellar luminosities and distances constitute an important part of the present-day program of observation, and the classification according to temperature of spectral lines in laboratory sources is continuingly fruitful both in its direct results and in its applications. In physical studies of the spectra of the cooler stars, reference to the spectrum of sunspots for comparison purposes has become almost automatic because of the relatively complete knowledge of spots gained in the course of the early investigation. Finally, the emphasis laid upon the conception of luminosity and its relationship to spectral characteristics has had a profound influence upon theories of stellar radiation and the life history of stars.

In this brief outline of a relatively simple research we have passed from a sunspot to the interior of the atom, thence to the atmosphere of the sun and stars, and so to problems of stellar constitution and development. That the behavior of certain lines in the spectrum of a sunspot should have any bearing upon the determination of the distance of a star seems at first sight almost inconceivable, but the successive steps in the relationship are logical and by no means complicated. The investigation forms an excellent illustration of the innumerable interconnections of the problems of physical science, of the simplicity and directness with which they may be approached, and of the effectiveness of bringing to their solution varying interests and a diversity of points of view.

THE BRIGHTNESS OF STARS: STANDARDS OF MAGNITUDE AND COLOR

FREDERICK H. SEARES

Assistant Director, Mount Wilson Observatory
The Carnegie Institution of Washington

I

To the eye of the observer the brightness of a star is nothing sensational; and yet, first and last, it brings to us a surprising amount of knowledge of the heavens. For example, it tells both the size of a star and where the star is—not without other information, to be sure; but with other information at hand, it is the one item needed to complete the answers to many questions about the structure of the universe. Moreover, it is the only clue we have to the distances of very remote objects; other methods of measuring distances break down before we have passed the doorstep to our own immediate region of space. Without measurements based on the brightness of stars, we should know nothing of the extent of our stellar system or of the distances of those other systems which, somewhat anomalously, we still speak of as extragalactic nebulæ.

The term "brightness" has behind it a long history that need not be recounted here. Originally it stood for the intensity of the visual sensation produced by the radiation emitted by a star, that is, the intensity of a physiological response to an external physical stimulus. Now it means merely a simple mathematical function of I, the intensity of the radiation received from the star, namely,

$$m = -2.5 \log I + \text{const.} \tag{1}$$

This expression defines the unit of brightness, the stellar magnitude.

Since the days of Hipparchus the brightness of stars has been expressed in magnitudes (m). The gap between the physiological concept of brightness and the purely physical concept implied by the foregoing definition is bridged by the omitted history.[1] Numerically, however, the ancient estimates are essentially the same as the magnitudes derived from equation (1). The modern procedure supplies the precision of a rigorous mathematical definition and a generalization which permits measurements of brightness, not only with the eye, but also by means of a photographic plate, a photoelectric cell, or any other device responsive to radiation.

The simplicity of equation (1) is illusory, however. The symbol I conceals a whole series of difficulties which begin to emerge into the light when this quantity is set down in symbolic form, thus:

$$I = \int_0^\infty \frac{\pi r^2}{\rho^2} \mathrm{E}(\lambda,\mathcal{T})\mathrm{A}(\lambda,h,z)\mathrm{T}(\lambda,d,I)\mathrm{R}(\lambda,I)\, d\lambda \ . \qquad (2)$$

With proper care, the actual measurement of I can be managed. The question is rather, What have we measured? To answer this question, the meaning of equation (2) must be made clear.

The radiation of wave length λ to $\lambda + d\lambda$ emitted by a star of temperature $\mathcal{T}$ appears in the equation under the symbol $\mathrm{E}\ (\lambda, \mathcal{T})d\lambda$, here expressed as a surface intensity (energy flux per unit solid angle summed over unit area of the radiating surface). Multiplied by the area of the star's disk, πr^2, and divided by the square of the distance of the star, ρ^2, this quantity becomes the total energy flux received per unit area at a point just outside the earth's atmosphere. To avoid further complication of the general photometric problem, it is here assumed that no energy has been lost by interstellar absorption during transmission of the energy to the earth.

The energy flux entering the earth's atmosphere is diminished by absorption in the atmosphere itself and in the telescope, differently for each wave length λ. The atmospheric extinction, in the simplest case, depends on h the altitude of the observing station above sea level and on z the zenith distance of the star. We therefore find in equation (2) the reduction factor $\mathrm{A}(\lambda, h, z)$. The character of the stellar image formed by the telescope depends on d its distance from the optical axis and frequently on the intensity I itself. The reduction factor for the telescope therefore has the form $\mathrm{T}(\lambda, d, I)$.

Finally, the energy that escapes the filtering action of the atmosphere and the telescope meets some form of receiving apparatus, which usually is highly selective in the wave lengths to which it responds. The receiver may be a human eye, a photographic plate, a photoelectric cell, a thermocouple, a radiometer, or any other form of physical apparatus sensitive to the impact of a stream of energy. In practice, different kinds of photographic plates may be used, each sensitive to radiation from a different spectral region, and, in combination with all these varied types of receiver, filters of various sorts which transmit only the energy falling within selected intervals of wave length. The response of the receiver to energy of a given wave length λ—the sensitivity curve of the receiver—differs to some extent for every eye and every brand of photographic plate. For photographic observations the response also depends on the temperature and humidity at the time of exposure and on the development of the plate. Moreover, the intensity I itself may here again influence the response; both eyes and photographic plates are subject to the Purkinje phenomenon. For example, suppose the responses to two different wave lengths to be the same; if then both energy intensities be changed in the same ratio—doubled or halved, say—the responses, in general, will no longer be equal. Only with the thermocouple, radiometer, and bolometer do we meet with a little welcome simplicity; these devices respond equally to energy of all wave lengths, but their general sluggishness restricts their use to bright stars. Since the circumstances affecting the receiver are so numerous, we merely state their existence, and use in equation (2) only the symbol $R(\lambda,I)$ to indicate the functional response of the receiver to energy of the given wave length.

---

[1] F. H. Seares, Magnitudes Again, *Science*, vol. 87, p. 1, 1938.

In delivering the quantity actually observed, the receiver lumps all wave lengths together and presents a sum total, I, expressed by the integral in equation (2) which extends over the entire range in λ from zero to infinity. It should now be clear why it is difficult to know what, in any given case, has actually been measured. What, for example, is the relation of the observed I, commonly expressed in stellar magnitudes by means of equation (1), to the physical properties of the star? In examining equation (2) with this question in mind, we are struck, first, by the fact that only the energy within a relatively narrow interval of wave length is included in our measures; and second, that this interval varies greatly with the receiver and to a high degree is arbitrary. When the eye serves as the receiver, we deal with the ordinary visual magnitude; but we have a whole series of other magnitudes, corresponding to other types of receiver; and usually the energy measured is only a small fraction of the total received from the star.[2] The inference to be drawn from these circumstances is the necessity for careful definition and standardization of the whole system of magnitudes. In fact, emphasis of this necessity for definition and standardization is the chief purpose of this rather detailed approach to the subject.

II

The definitions underlying our conventional series of magnitudes —notably visual magnitudes and photographic magnitudes—were adopted almost automatically. One was fixed by the spectral sensitivity of the average eye, the other by that of the ordinary silver-bromide plate. Photovisual magnitudes also dropped naturally into place; an isochromatic plate exposed behind a yellow filter was found to be practically equivalent in responsiveness to the average eye. These magnitudes, presenting all the advantages of results obtained by photography, have come more and more to be used as a substitute for ordinary visual magnitudes.

At first little attention was paid to the differences in the responsiveness of individual eyes and of different brands of plates, or indeed to the influence of the telescope. A telescope was a telescope, and if differences in color response made their appearance, the observer and the plate were to that extent abnormal and there was not much to be done about it.

The relationship of all these quantities to the total radiation of the star remained obscure; but it was soon recognized that the magnitude difference, photographic minus visual (or photovisual), was definitely related to the color of the star, and hence to its temperature. Obviously this difference could be used as a measure of color or temperature, and it was therefore appropriately named "color index." The introduction of this concept connected the magnitude system with an important physical property of stars; it also provided a means of intercomparing different series of magnitudes of the same stars measured with different telescopes and different plates. When the necessity for accurate standardization had become apparent through closer attention to the details of equation (2), it was found that in most cases, or to a first approximation at least, the differences in magnitudes of the same kind arising from small differences in the instrumental equipment—from variations in color system, in other words—were proportional to the color indices of the stars. Here then was a simple and practicable method of reducing magnitudes to a standard system of color.

---

[2] F. H. Seares, Magnitudes Again, *Science*, vol. 87, p. 1, 1938.

This result is of great practical importance, for clearly the observations themselves cannot be completely standardized. Atmospheric extinction can be allowed for; but the observer must use the telescope at hand, and generally also the plates easily obtainable. Some means for securing homogeneity in the final results is therefore essential; and with that provided, we lack only a definition or convention as to what the standard color system shall be. The adopted color system for photographic and photovisual magnitudes happens to be that defined by the silvered reflector and certain brands of plates, the old Seed 27 emulsion, and the Cramer Instantaneous Iso used with a certain yellow filter. Practically, it is defined by the actual photographic and photovisual magnitudes adopted by international agreement for a certain group of stars near the north pole. The origin of these standards is involved with the establishment of the scale of magnitudes.

III

The approach to the problems of photometry by way of the physical factors involved in equation (2) is not the chronological order of development. In the beginning the observer was so busy with the practical difficulties of measuring radiation and establishing a reliable scale of magnitudes that he did not inquire closely as to what he had measured; he was concerned mostly with equation (1), and in particular with the satisfaction of the condition imposed by the coefficient 2.5, which defines the magnitude scale.

The measurement of radiation, as compared with the measurement of things that can be touched and handled, is a difficult matter, and the precision attained is never high. Moreover, the range in intensity which must be covered is enormous. The interval between the brightest stars and the faintest reached by powerful modern telescopes is roughly 20 magnitudes; the corresponding range in intensity, in accordance with equation (1), is 100,000,000 to 1. To establish the scale over such an interval as this with any approach to accuracy requires much time and much diversification of method. If, however, the brightness of any group of stars is known over a long range in magnitude, the magnitudes of other stars can be found with no great difficulty. This circumstance determines the procedure usually followed in measurements of brightness: first, establish a series of standard magnitudes in some conveniently chosen location in the sky, say at the north pole; then determine the magnitudes of other stars, as required, by comparisons with the polar region.

This method seems first to have been suggested by E. C. Pickering; and Pickering's pioneer work at the Harvard Observatory in setting up standards of magnitude, supplemented by that of Miss Henrietta Leavitt,[3] did much to systematize the measurements of stellar brightness. The standards near the pole chosen by Pickering were selected a few at a time, until by 1912 the list included the 96 stars now known everywhere as the North Polar Sequence.

Work on this same problem of establishing a reliable scale was begun at Mount Wilson in 1910. Kapteyn, then a research associate of the Mount Wilson Observatory, was anxious that standard sequences of magnitudes be established in each of the 139 Selected Areas that could be observed from Mount Wilson. As a preliminary to that investigation, which resulted finally in the preparation of the *Mount Wilson Catalogue of Selected Areas*, various methods of establishing the scale were studied in detail, and to test these methods both photographic and photovisual magnitudes were derived for each of the stars of the Polar Sequence.

In principle all methods of establishing the scale require some kind of measurement of the intensity of starlight. Practically, this may be accomplished by making two equal exposures to the field of stars on the same photographic plate, one with the full aperture of the telescope, the other with the aperture reduced by a known amount, say the equivalent of 3 magnitudes, by means of a diaphragm, a wire gauze screen, or some similar device. If then the full-aperture image of a faint star equals the reduced-intensity image of a brighter star, the actual difference in brightness of the two stars will be 3 magnitudes. A simple extension of this principle gives values for the relative magnitudes of all the stars on the plate. The zero point, defined by the constant in equation (1), is arbitrary and must be fixed in accordance with some predetermined convention.

---

[3] See especially *Harvard Circulars*, Nos. 108, 125, 150, 160, 170; *Harvard Annals*, vol. 71, no. 3.

When examined in the light of equation (2), the results thus found present certain noteworthy features. First, a correction for atmospheric extinction must be applied, commonly included with the zero-point correction and usually so determined as to give what would have been obtained had the stars been observed in the zenith. Second, corrections for distance of the images from the optical axis must also be made, a correction which often depends on the brightness of the star. In the ideal case of perfect observations, the resulting magnitudes and the intensities of the stars will then satisfy equation (1). To minimize the influence of accidental errors on the scale thus established, the measurements must be repeated many times. To avoid the appearance of systematic errors, many precautions must be taken. In particular it is important to reduce the intensity by many different devices and by varied amounts. Third, for observations made with the same telescope and the same brand of plate, all similarly developed, the results should be homogeneous in the sense that they represent for each star a similar integration of intensity over the same interval of wave length. In other words, the results are homogeneous in color system. Finally, if both photographic and photovisual magnitudes have been observed, their differences afford a measure of the colors of the individual stars which can easily be calibrated in terms of spectral type or temperature when these data are available. The relation of the integrated I to the total radiation of the star remains undetermined, however, and, when needed, must be found by an independent investigation.

The results obtained by various methods used with the 60-inch reflector at Mount Wilson were satisfactorily accordant but deviated considerably from the scale established at Harvard. Part of this discrepancy was accounted for by differences in color system. The Harvard measures had been made with many different instruments and were not reduced to a homogeneous color system until the Mount Wilson color indices for the fainter stars became available. The Mount Wilson magnitudes and colors and a rediscussion of the Harvard data were published in 1915.[4] To clear up the remaining scale divergence of 6 per cent affecting the stars brighter than the tenth magnitude, additional measures were made both at Harvard and at Mount Wilson[5] and new investigations were undertaken at several other observatories. The result was a general confirmation of the Mount Wilson measures.

---

[4] F. H. Seares, *Mt. Wilson Contr.*, Nos. 97, 98; *Astrophys. Jour.*, vol. 41, pp. 206, 259, 1915.

IV

With the organization of the International Astronomical Union in 1919, the question of suitable standards, which was taken in hand by the Commission on Stellar Photometry, became a cooperative matter. All the available data on the Polar Sequence were reduced to the color system of the Mount Wilson measures made with the reflector and combined into mean values, which were presented to the Union at the Rome meeting of 1922 and there adopted as international standards.[6] The photographic standards thus established for stars brighter than the thirteenth magnitude depend on observations made at six different observatories with ten or a dozen different telescopes, but the fainter stars were observed only at Mount Wilson. The photovisual standards were based wholly on Mount Wilson measures, but their consistency with the photographic scale was checked by measurements of color made by the method of exposure ratios,[7] which is entirely independent of determinations of magnitude. Later, the photovisual scale to the twelfth magnitude was confirmed by visual measurements made at Potsdam.[8]

A further check of great importance on the accuracy of the standard photographic scale was obtained in 1935 and 1936 through the cooperation of Stebbins and Whitford, who measured both the magnitudes and the colors of thirty of the brighter stars of the Polar Sequence with a photoelectric cell attached to the 60-inch reflector.[9] The comparison of these results with the international standards gave considerable trouble owing to the rather complicated relationship of the photoelectric colors to the standard color indices. A reduction of the photoelectric magnitudes to the international photographic system made by Seares[10] shows, however, excellent agreement in scale over an interval of nine magnitudes, except for four bright stars, Nos. 1 to 4 of the Polar Sequence series. These four stars are all early A in spectral type, and hence all affected by strong hydrogen absorption in the stellar atmospheres. Owing to differences in the sensitivity curves for the photographic plate and the photoelectric cell, the hydrogen absorption should modify the photographic magnitudes of these stars but not the photoelectric values. The difference is of the right sign and apparently of the right order to account for the observed discrepancy of 0.1 magnitude, but at present the matter cannot be decided with certainty.

---

[5] F. H. Seares and M. L. Humason, *Mt. Wilson Contr.*, Nos. 234, 235; *Astrophys. Jour.*, vol. 56, pp. 84, 97, 1922.

[6] *Trans. Internat. Astr. Union*, vol. 1, p. 69, 1922.

[7] F. H. Seares, *Mt. Wilson Contr.*, No. 235; *Astrophys. Jour.*, vol. 56, p. 97, 1922.

[8] F. H. Seares, *Mt. Wilson Contr.*, No. 431; *Astrophys. Jour.*, vol. 74, p. 131, 1931.

[9] *Mt. Wilson Contr.*, No. 586; *Astrophys. Jour.*, vol. 87, p. 237, 1938.

V

The list of international standards includes values as faint as 20.1 photographic and 17.4 photovisual, but below 19.0 and 15.5, respectively, the results are of lower precision. As already remarked at the end of section III, these values of the photographic and photovisual magnitudes are themselves a definition of the international color system. In practice it is easy to calibrate any given instrumental equipment in terms of this system by observing the standards with the equipment in question. If the magnitudes of blue stars thus observed are made to agree in zero point with the standards, the results for red stars will in general disagree. The differences will be nearly, if not exactly, proportional to the color index, and the relationship thus found provides the necessary corrections for reduction to the standard system.

Although the precision of the international standards seems to be satisfactory, the Polar Sequence, as a whole, leaves something to be desired in convenience of usage. The number of standards, among the brighter stars especially, is too small. A more complete representation of color indices is also needed. With only the stars of the Polar Sequence as a basis, it is not always easy to transfer the standard scale and color system to other points of the sky. Particularly is this true if the observations are to be made with a short-focus, large-field camera. The corrections for distance of the stars from the optical axis and for reduction to the international color system required by observations made with these instruments are complicated and cannot be accurately determined unless numerous stars of known magnitude are available. Further, although the general accessibility of standards situated very close to the pole is an obvious advantage, such a location carries with it an observational complication if the required exposures are long. Guiding of the telescope during exposures centered exactly on the pole is notoriously difficult, and the resulting images are subject to slight deformations which do not appear on plates centered only 3° or 4° distant from the pole. These differences are frequently a source of systematic error in the magnitude scale derived for other regions of the sky by means of comparisons with the pole.

---

[10] *Mt. Wilson Contr.*, No. 587; *Astrophys. Jour.*, vol. 87, p. 257, 1938.

These deficiencies in the Polar Sequence were the origin of a current investigation made at Mount Wilson with the cooperation of Dr. Frank E. Ross of the Yerkes Observatory, which is now approaching completion.[11] Some 2000 stars brighter than about 11.5 on the photographic scale, located between 80° north declination and the pole, have been extensively observed with several short-focus cameras for the determination of both photographic and photovisual magnitudes on the international system. In addition, all the earlier observations of these stars, visual, photovisual, and photographic, have been reduced to the international system and combined with the recent results. The mean values provide supplementary standards, frequently comparable in precision with the Polar Sequence standards themselves, which will greatly simplify future work on the brighter stars. Incidentally, in determining the additional magnitudes, each of the brighter Polar Sequence stars has been measured many times relative to other stars of the Sequence. Theoretically at least, these measurements provide data for decreasing the effect of accidental errors on the original Polar Sequence standards. A readjustment of the magnitudes based on these data has accordingly been made; at the same time certain recent measurements of the colors of some of the Sequence stars by Malmquist and by de Sitter have also been included in the adjustment.[12]

---

[11] Seares, Ross, and Joyner, Magnitudes and Colors of Stars, North of +80°, mimeograph edition, Pasadena, 1935. A preliminary publication giving photographic magnitudes for 2269 stars, and provisional photovisual results for about 400 of the brighter stars of the list.

[12] F. H. Seares, *Mt. Wilson Contr.*, No. 587; *Astrophys. Jour.*, vol. 87, p. 257, 1938.

We have, therefore, finally, a list of fundamental standards, both photographic and photovisual, referred to what appears to be a homogeneous color system, which depends on all the available measures made by photographic methods. The completion of the work on stars north of +80° declination will supply a large number of supplementary standards for brighter stars which will greatly strengthen the system of magnitudes and colors and simplify its transfer to other parts of the sky.

In this connection the *Mount Wilson Catalogue of Selected Areas* already referred to should perhaps be mentioned again as supplying secondary standards of photographic magnitude for 139 fields uniformly distributed over the sky north of −15° declination. These results, all on the international system, have been shown by repeated tests to be uniform in zero point and reliable in scale between magnitudes 13 and 17.5 or 18.0.

As a whole, the various investigations underlying the Polar Standards are a good example of successful cooperation in an undertaking in which cooperative effort is especially advantageous. The observations are difficult and peculiarly subject to systematic errors. The generally excellent agreement of results obtained by many observers, with different instruments and by a wide variety of methods, encourages the hope that the problem of magnitudes has been placed on a sound foundation.

The opening paragraph of this brief account of stellar brightness has suggested the significant part played by the magnitudes of stars in the problems of modern astronomy and may have led to the expectation that some account of the contributions to our knowledge made by measurements of brightness would appear somewhere in the discussion. The purpose here has been, however, not to recite the extraordinary and far-reaching results thus obtained, but instead to display certain characteristics of the concept of magnitude itself, which today serves as an instrument of research. This aspect of the matter is important. If a tool is to be used intelligently, it must be understood; and in the past there has been much misunderstanding. It may be regarded as a matter of emphasis, therefore, that this account stops where it does.

MAGNETISM AND THE ATOMIC NUCLEUS

M. A. TUVE, L. R. HAFSTAD, AND N. P. HEYDENBURG
Department of Terrestrial Magnetism
The Carnegie Institution of Washington

INTRODUCTION

FOR some years the Department of Terrestrial Magnetism has been carrying forward a laboratory program specifically directed toward the atomic nucleus and studies of the interactions of the primary particles of matter. This pioneer project for the use of extremely high voltages for examining the intimate structure of matter, initiated in 1926 with the help of our colleague Dr. Breit and now duplicated in so many laboratories of the world, was planned to develop the most fundamental investigations which seemed then possible regarding the basic nature of magnetism itself. It concerned the laws governing the interactions with each other of the simplest magnetic bodies or of single magnetic particles. At that time it was just being discovered—by work on the spinning electron and on nuclear magnetic moments—and it has since been amply demonstrated, that one of the three or four known attributes of the primary particles of all matter is that they all possess magnetic moment or intrinsic magnetization.

As far as the magnetic field of the Earth is concerned, such studies as these may have small *direct* significance but large *indirect* significance, as through them fundamental understanding of the nature of magnetism itself will be advanced. One really outstanding puzzle in terrestrial magnetism, still unexplained by known laws of physics, is the existence of the enormous permanent magnetic field of the Earth. The numerous important problems relating to variations of the Earth's magnetic field—variations only infrequently as large as 1 per cent of the total field—can now be explained to some extent even quantitatively. This is especially true since the results of exploring the upper atmosphere by radio waves, as originated at the Department in 1925, have become available. However, the large steady magnetic field of the Earth may not be explained by a magnetization of the thin, solid crust. There seems no longer any doubt that the interior of the Earth is at a temperature much too high to allow ferromagnetic behavior despite a possible shift of the Curie-point with pressure (experiments seeking this effect by the Department in cooperation with the Geophysical Laboratory have given negative results). Magnetization due to the Earth's rotation had been disposed of by the investigations of Barnett at our Department and others. Similarly, accumulated evidence indicates that the electrical conductivity of the interior of the Earth could not possibly be large enough to permit the long-continued existence of a circulating current within the Earth, quite aside from the difficulty of accounting for the origin of such a current. Thus by 1926 it was realized that, in addition to the importance of such fundamental information for its own sake, it was desirable to enlarge knowledge of the basic nature of magnetism itself through investigation of the detailed laws governing the interaction with one another of the magnetic particles composing all matter. It is possible that the large and unexplained magnetic fields of the Earth and of the Sun, each related in the same way to their directions of rotation, may arise from some unknown characteristic of these primary laws.

The results to date of this program, although they have uncovered other facts of great significance and have measured the deepest forces within the atom, have given no clue as to why the Earth and the Sun have such large magnetic fields. Perhaps some obscure atomic effect of the extremely high pressure at the centers of these large bodies may be the cause of the magnetization but as yet there is no indication of any way to explain it. It is reasonable to believe, however, that the permanent magnetic field of the Earth is a problem of complexity for which there is as yet no basis for calculation, rather than a problem of concept or fundamental idea. Almost certainly it will not require a revision of the well-authenticated laws of electromagnetic action, although some unexpected phenomenon, such as to constitute in effect a large electrical current inside the Earth, may ultimately be the explanation.

The Forces Which Govern the Atomic Nucleus

What is meant when we speak of the forces acting deep within the atom? What may be the validity of our ideas and statements about anything so far removed from our direct perception and experience? The minute dimensions of the central heavy cores or nuclei of atoms of the various chemical elements have been measured directly by shooting high-speed particles (for example, nuclei of hydrogen or helium atoms) at thin films of material and observing that hundreds of millions of these particles pass straight on through for every one that is appreciably deflected. Solid bodies are literally full of holes; all atoms are comprised largely of empty space. Atoms have their mass or weight almost entirely concentrated in central nuclei, a hundred thousand times smaller than the atoms themselves. The relatively large dimensions of atoms are simply the dimensions of the atmospheres of light-weight electrons which surround their nuclei. Does even the *idea* of a force have any meaning in a region of space ten million million times smaller than our own thumbs and fingers? Defining what is meant by force, the question then becomes one as to whether the ideas of force and energy and mass and momentum, defined in accordance with our experience with larger things, are *useful* ideas in a discussion of these small regions.

We assume (1) that mass or inertia is a basic property of the small particles of which matter clearly appears to be built and (2) that mass, measured in accordance with our definitions, cannot be created or destroyed. Then to what extent may these ideas, together with the familiar concept of force, be useful to predict the behavior of the atomic or sub-atomic particles under given new conditions from our observations of their behavior under other different conditions? The answer is that force and mass are useful to a remarkable degree. Nearly all the information available about the inside of the atom and about the interactions between atoms may be organized in terms of force, mass, energy, and momentum, similar to their application in the physics of things of ordinary size, although assuredly there still remain certain puzzles.

The Primary Physical Forces

The main ideas may be outlined briefly. Physical science is now so comprehensive that all interactions between material bodies may be described in terms of just three different kinds of fundamental forces. All the phenomena of actions and reactions in the world of physical things, from astronomy down through mechanics and heat and radio and light, and through chemistry and atomic structure into the atomic nucleus itself, are explicitly expressions of gravitational force, electromagnetic force, and the force which governs the atomic nucleus. Perhaps these may ultimately be reduced to different aspects of some one great, all-pervading and all-inclusive type of force. However, the description of the whole universe in terms of just three kinds of force is a marked simplification of the apparent near-chaos.

If the analysis is fine-grained enough to deal with the atom and its parts, two elements must be introduced which are not necessary when dealing with things or systems of ordinary size. It is well known that in dealing with gravitational forces the ordinary analysis had to be modified by the introduction of the concepts of the relativity theory when the analysis was extended to extremely large dimensions. Similarly, in the direction of extremely small dimensions dealing with atoms and atomic parts, it is necessary to make two modifications, namely, to introduce an element of discreteness or discontinuity, as indicated by the name "quantum-theory," and to take account of what is called the phenomenon of exchange. These two modifications evolve from recognition of the essential limitations of our concepts of the real nature of matter—the nature of particles and the nature of waves—when these ideas are carried down to exaggeratedly microscopic dimensions. This is just a repetition of the old dilemma of particle-theory versus wave-theory, well illustrated by the history of theories regarding the nature of light. Newton explained light in terms of particles; the discovery of phenomena of interference a century ago required light to be wave-motion; today it is recognized that there is an essential duality about light: it has both wave-characteristics and particle-characteristics. Similarly, the understanding of atoms involves duality of the properties of particles and waves, expressed in the modern form of the quantum-theory called the wave-mechanics.

Another point to be remembered is the distinction made of a problem or a difficulty as "fundamental." A fundamental difficulty means a trouble which is inherent in basic concepts, the failure of the idea itself, as distinguished from a difficulty of complexity, the failure of mathematical or experimental abilities to meet the needs of a problem. It is impossible, for example, to write a detailed mathematical solution describing the motions of a hundred particles all interacting simultaneously with one another; even the classical three-body problem can be solved only with restrictions. However, the interaction of two bodies represents no such difficulties of complexity; should the formulas fail to describe a two-body atomic problem correctly, for instance, the trouble is with the ideas or concepts themselves. It is such formulation of a fundamental problem which centers emphasis in nuclear research at the Department of Terrestrial Magnetism so definitely on essentially simple problems—for example, the interaction of a proton (the nucleus of an ordinary atom of hydrogen) with another proton, or of a proton with a neutron (a particle of the same mass but without the positive charge of one electron which is carried by the proton). Protons and neutrons are the primary heavy particles of which the nuclei of all atoms are built.

Indirect Information Concerning Nuclear Forces

Mass-defects. Prout's hundred-year-old hypothesis that all atoms are built of units essentially the same as hydrogen atoms was placed on a firm experimental basis by the experiments of J. J. Thomson some 20 years ago. Thomson showed that many (now most) of the chemical elements are each in turn mixtures of different species of atoms—separate species of the same chemical element—differing from each other by one unit of mass or weight. This unit is approximately the same as the mass of one hydrogen atom, although significantly smaller. Since the discovery of the neutron in 1932, the view has been accepted that all nuclei are built of neutrons and protons. The nucleus of a hydrogen atom—the proton—possesses an electrical charge exactly equal and opposite to that of a negative electron, and is approximately 2000 times as heavy as an electron. Ultimately it may be considered that nuclei are built up of neutrons and positrons (positive electrons), a proton in the nucleus then being considered as dissociated into a neutron and a positron. This view does not, at present, lend itself favorably to any theoretical formulation; however, it may well be that protons lose their identity in the close confines of the nucleus. Obviously nuclear structure is inherently complex and correspondingly difficult to formulate, since all nuclei have dimensions not very much larger than the protons and neutrons of which they are built, assuming as the "sizes" of these particles the distances of separation at which large non-electrical forces become abruptly evident.

When protons and neutrons are bound together to form atomic nuclei an amount of energy is given off, in the form of radiation or otherwise, which is called the energy of binding. Since energy and mass are equivalent properties in accordance with special (restricted) relativity, this binding-energy is evidenced as a loss of mass in the compound nucleus, the mass of which is smaller than the masses of the protons and neutrons which entered its structure by an amount which is an exact measure of the binding-energy, the energy that would have to be supplied from an outside source to separate the nucleus again into its component parts. This "mass-defect" gives an exact measure of the end-result of the actions of the nuclear forces. Mass-defects throughout the atomic table have been carefully measured. They show one outstanding property, namely, that the energy of binding per particle is approximately constant throughout the atomic table, being roughly eight million electron-volts for each proton or neutron which enters the nuclear structure. This characteristic is referred to as "saturation"; any hypothetical kinds of nuclear forces must therefore exhibit this property of saturation.

Nuclear transformations. A second source of information regarding the nuclear forces is found in the numerous data relating to nuclear transformations, or atomic transmutations ("atom smashing"), accumulated since Rutherford's first demonstration of artificial transmutation in 1919 and particularly since Cockcroft and Walton, in Rutherford's laboratory, first accomplished the same feat in 1932 using as bombarding particles protons which were accelerated by a high voltage applied to a vacuum-tube. Incidentally this latter was one of the possibilities envisioned when the Department's high-voltage program was begun in 1926. The data on atomic transmutations give primarily a confirmation of the data on mass-defects. When a transmutation occurs with the evolution of energy (resulting in a loss of mass, a higher binding-energy) this energy appears frequently as kinetic energy of the resultant system of particles (one or several particles are emitted from the complex temporary system formed by the projectile and the target nucleus, leaving a residual nucleus which recoils as the particle is shot out). Measures of these large reaction-energies confirm the mass-defect data but the measurements on nuclear transformations give additional important information. They determine the probabilities of various kinds of reactions—so to speak, the rate of activity of the nuclear forces and the degree of stability of various combinations or states of motion of the nuclear particles. This last remark is especially true because many of the products of reaction of nuclear transformations have only a limited stability, even though they are nuclei of the ordinary chemical elements and behave with perfect propriety in all chemical reactions. If a proton or a neutron is added to a nucleus to form a new nucleus which is one unit heavier or one unit lighter than the ordinary stable nuclei of the same chemical element as found naturally, it is usually observed that the extra-light or extra-heavy nucleus "blows up" later. These nuclei explode in the same way as do the nuclei of the well-known radioactive elements at the heavy end of the atomic table; they possess a certain fixed probability which leads to the spontaneous emission of a negative or a positive electron during a given time. Accordingly, at the end of a period which varies from a fraction of a second to months or years, depending on the particular nuclear species, half of the unstable nuclei present at the beginning of the period have transformed themselves into a stable species of nucleus belonging to the adjacent element in the atomic table. This phenomenon is called artificial radioactivity, discovered by Irene Curie and F. Joliot in 1934.

High-voltage equipment and technique at the Department of Terrestrial Magnetism. Because some nuclear transformations are millions and even billions of times more probable than others, the problem of unraveling the meaning of simple observations on the particles emitted by given targets under bombardment is not a simple one, since no targets are ever really chemically pure. It became clear early in the Department's investigations that reliable conclusions could be obtained only by quantitative observations of the highest order. The quantitative yields of various reactions vary quite differently with voltage, and accordingly the most analytical technique is one in which the voltage or energy of the bombarding particles is under accurate control and is subject to accurate measurement. This includes, of course, the requirement that the bombarding beam must be very nearly homogeneous in kind and energy.

The high-voltage source (plate 2) used is an old-fashioned electrostatic generator charged by a belt, a device re-invented at an opportune time (1931) by Dr. R. J. Van de Graaff, then at Princeton University. (The idea of belts for conveying the charge was proposed as long ago as 1870 by Lord Kelvin and by Righi, who constructed somewhat similar machines of small size at that time.) The high-voltage electrode is simply a ball of large diameter, following the standard electrical practice of preventing corona-losses by using electrodes of suitably large dimensions and small curvature (similar spherical "corona-caps" were used in all the experiments with Tesla coils and high-voltage tubes at the Department during 1926 to 1930). At the top of the vertical glass vacuum-tube, inside the 6-foot ball, is a small metal tube filled with a hydrogen (or heavy hydrogen) discharge at low pressure and low voltage (supplied by a small 110-volt generator driven by an auxiliary insulating belt). Protons or deuterons or helium ions (nuclei of light or heavy hydrogen or of helium, produced as ions in the low-voltage discharge) pass through a small canal from this discharge into the high-voltage tube. Here they are focused as a beam, which is accelerated down the axis of the tube to strike a target placed in a "grounded" extension of the vacuum-tube which projects downward into the observing-room beneath the high-voltage generator. Before striking the target where the transmutations are to be produced, this beam of high-speed particles passes through a strong magnetic field, which is so adjusted as to allow only ions of the desired kind to impinge on the target—for example, protons (mass 1) or deuterons (mass 2) or helium ions (mass 4). This equipment produces a bombarding beam of great homogeneity in kind and energy (energy-spread under 1 per cent) at any desired voltage from 200,000 to about 1,150,000 volts. The voltage (energy of the particles) is measured accurately to 1 per cent and continuously by a voltmeter-resistor (10,000 megohms) enclosed in a large Textolite tube, shown projecting at an angle downward from the high-voltage "ball."

Direct Observations on the Nuclear Forces

One of the aims of the high-voltage program set in 1926 proved, in its realization ten years later, of great importance in formulating definite ideas regarding the atomic nucleus. This was to obtain measurements on the collisions of protons with protons to determine any deviations from the predictions based on the ordinary inverse-square law of repulsion between like electric charges. Such deviations would be direct experimental information concerning the forces which had been assumed for many years to hold together the primary particles constituting the atomic nuclei—the nuclear forces.

Neutron-proton scattering. The first direct observations of the nuclear forces were observations of the scattering of neutrons by matter. There are three forces active in the structure of the nucleus if nuclei are built up of neutrons and protons, namely, (1) proton-neutron forces, (2) proton-proton forces, and (3) neutron-neutron forces. Observation of (1) by the scattering of neutrons by hydrogen gave direct evidence regarding the proton-neutron interaction. Measurements at the Department of the proton-proton interaction evidenced the second and thereby defined the third. This led, through theoretical analysis of the experimental curves by Dr. Breit, to the simple conclusion that all three types of interaction are nearly, although not exactly, the same.

Proton-proton scattering. Protons are accelerated to a specified velocity (energy) by the high-voltage tube and pass through a "window" of aluminum foil into a scattering chamber (figure 1) which is filled with pure hydrogen gas at pressure of 12 mm. Diaphragms at the top define a 2-mm beam of protons passing through the gas of the chamber; protons scattered by "billiard-ball collision" from a small segment of this beam at the center of the chamber in such a direction as to pass through the set of diaphragms or slits fastened to the ionization-chamber are recorded as individual counts by a linear amplifier, similar to a radio amplifier. The loss of energy of the protons in the aluminum foil and in the gas is checked by measurements of the fluorine resonances using a target at the bottom of the lower tube, thus giving accurate knowledge of the velocity of the protons at the scattering volume. The latter is defined by the intersections of the lines extended from the two diaphragm-systems. The ionization-chamber with its diaphragm-system can be set at various angles to the primary beam.

The observations comprise measurements of the number of individual protons which are scattered through various angles (from 500 to 10,000 at each angle) by individual hydrogen nuclei (protons) in the small scattering volume when a known large number of protons of the primary beam passes through this volume. These

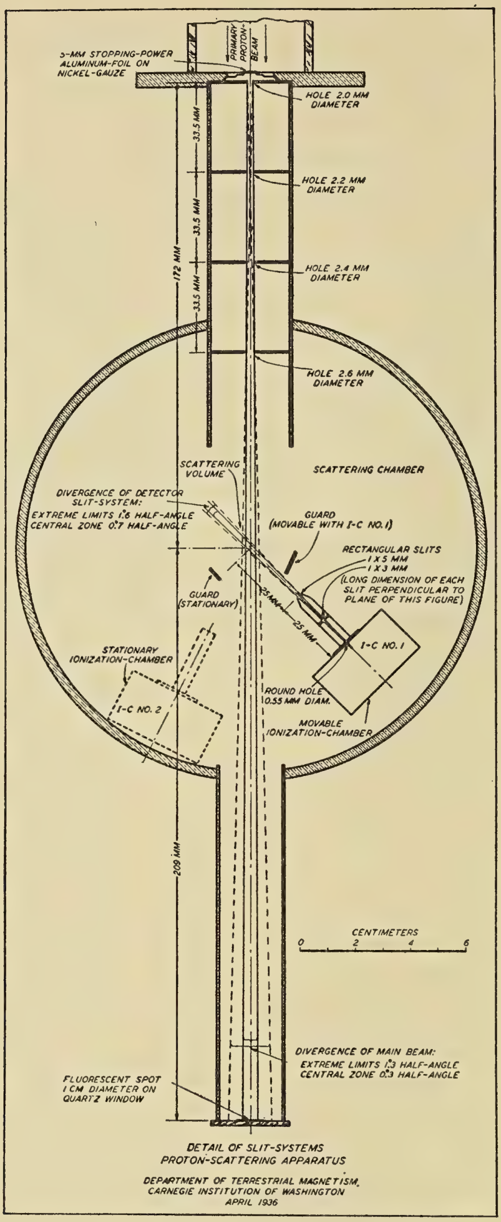

PRIMARY PROTON-BEAM
5-MM STOPPING-POWER ALUMINUM-FOIL ON NICKEL-GAUZE
HOLE 2.0 MM DIAMETER
33.5 MM
HOLE 2.2 MM DIAMETER
33.5 MM
172 MM
HOLE 2.4 MM DIAMETER
33.5 MM
HOLE 2.6 MM DIAMETER
SCATTERING VOLUME
SCATTERING CHAMBER
DIVERGENCE OF DETECTOR SLIT-SYSTEM:
EXTREME LIMITS 1°.6 HALF-ANGLE
CENTRAL ZONE 0°.7 HALF-ANGLE
GUARD (MOVABLE WITH I-C NO. I)
RECTANGULAR SLITS
1 X 5 MM
1 X 3 MM
(LONG DIMENSION OF EACH SLIT PERPENDICULAR TO PLANE OF THIS FIGURE)
GUARD (STATIONARY)
25 MM
25 MM
I-C NO. I
STATIONARY IONIZATION-CHAMBER
ROUND HOLE 0.55 MM DIAM.
I-C NO. 2
MOVABLE IONIZATION-CHAMBER
209 MM
CENTIMETERS
0
2
4
6
DIVERGENCE OF MAIN BEAM:
EXTREME LIMITS 1°.3 HALF-ANGLE
CENTRAL ZONE 0°.3 HALF-ANGLE
FLUORESCENT SPOT 1 CM DIAMETER ON QUARTZ WINDOW
DETAIL OF SLIT-SYSTEMS
PROTON-SCATTERING APPARATUS
DEPARTMENT OF TERRESTRIAL MAGNETISM,
CARNEGIE INSTITUTION OF WASHINGTON
APRIL 1936

FIG. 1. Diagram of the scattering chamber, used for measurements of the large nuclear force of attraction which was found to arise very abruptly when two protons (nuclei of hydrogen) approach very near to each other.

numbers are then compared with the numbers which are expected on the basis of the electrostatic repulsion between protons, after a formula by Mott. The ratio of observed number of counts to the number expected (for each given angle and voltage) gives a measure of the deviations arising from forces superposed on the Coulomb repulsion. Figure 2 shows a series of such ratios for angles from 20° to 45° and for energies from 670 to 867 kilovolts. If the inverse-square law were obeyed all these curves would coincide with the line drawn through the ratio of unity. Obviously besides the electrical repulsion between protons, some other force is acting and this force is much more important at 867 kilovolts and 45° than for lower voltages and angles. The protons come only slightly closer together in a collision at this voltage than they do at 670 kilovolts, so the additional force sets in quite abruptly when the protons come within a certain distance of each other.

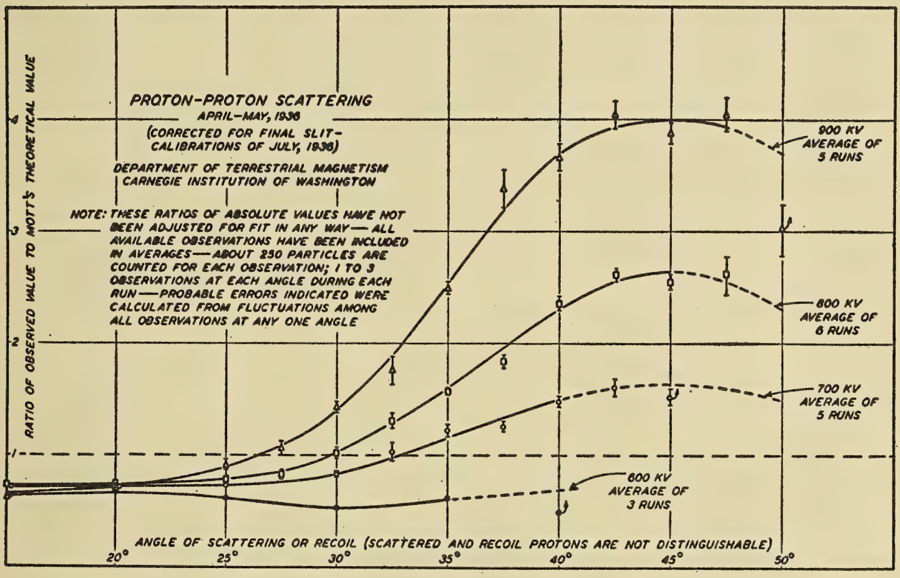


FIG. 2. Results of measurements on the angular scattering of protons by protons ("billiard-ball collisions") at energies from 600,000 to 900,000 volts. The large deviations from the values expected on the basis of the electrical repulsion between the two protons (each having a positive charge) shows the failure of the familiar inverse-square law of electrical repulsion due to the additional nuclear attraction between two protons at very close distances of separation.

That this "extra" force is an attraction and not an added repulsion, is evidenced (figure 3) by the ratios of observed counts to those predicted by the Mott formula for an angle of 45° and for energies from 200 to 900 kilovolts. At 400 kilovolts the observed scattering is nearly zero—a situation in which the repulsion due to the like electrical charges of the two protons is effectively just canceled by the attraction due to the nuclear force acting between them. At 200 kilovolts the scattering to 45° is caused primarily by the electrostatic repulsion, diminished somewhat by the nuclear attraction. Above 700 kilovolts the scattering at 45° results primarily from the nuclear attraction between the protons.

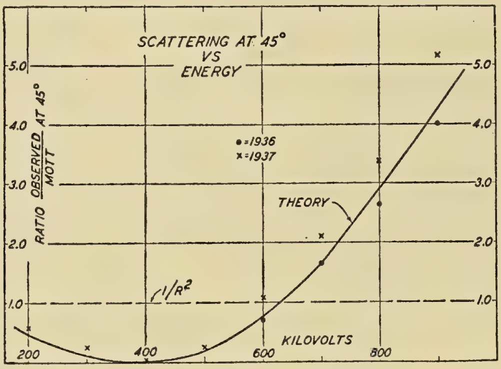


Fig. 3. Measurements of protons scattered to 45° for a range of energies from 200,000 to 900,000 volts. At low voltages the protons do not approach closely and the nuclear force has a small effect; at 900,000 volts the effect is large. The absence of scattering at 400,000 volts is due to the electrical repulsion being neutralized by the nuclear attraction. The curve marked "theory" refers to theory of nuclear forces, not the classical theory of electrical repulsion which predicts points on the dashed line marked "$1/R^2$."

There are two other important quantitative features of the experiments. One is the arrangement of a thin gold leaf mounted in the tube below the scattering chamber with a second counter set to observe the number of protons scattered by the gold. This gives a continuous measure of the proton-current passing through the scattering volume—in other words, a measure of the total number of primary protons involved in a given scattering measurement. The second feature is the method for "absolute" calibration (centimeters, grams, seconds) of the scale of the voltmeter by measuring the number of protons scattered by a given small pressure of spectroscopically pure argon introduced into the scattering chamber. Since a nucleus of argon has 18 times as great a positive charge as a proton, the great repulsion between the nuclei of argon and the protons prevents the latter from approaching within range of the nuclear forces around the nuclei of argon, and the scattering is exactly "classical," that is, it follows the formulas based on the inverse-square law of electrostatic repulsion. This has been tested experimentally for various angles and voltages using argon for the scattering gas (pure nitrogen and oxygen show very slight effects of the nuclear forces); the result showed it to be necessary to shift the original voltmeter-scale by somewhat less than 2 per cent to give correct absolute values.

Direct conclusions based on the empirical evidence. From these data, almost without reference to any form of theory regarding the nature of the nuclear forces, it is concluded that when two protons are brought within a certain very close distance from each other a "new" force makes its appearance, superposed on the ordinary electrostatic repulsion between the two like charges of the protons. This new force sets in abruptly and is an attraction, which rapidly overwhelms the large electrostatic repulsion and causes the protons to attract each other very strongly. This attraction is 10^{36} times the gravitational attraction between the two protons according to Newton's law and is about 10^{27} times the weight of either proton—a force of one kilogram acting between two protons, each weighing only 1/1,000,000,000,000,000,000,000,000 gram! It is not surprising that a high voltage is required to overcome the mutual electrostatic repulsion between charged particles which reaches such a magnitude before the nuclear attraction begins. Since the primary particles of matter are not mathematical points but have a certain diffuse (wave) character, "distance of approach" of the two protons for this range of nuclear attraction may not be defined with strict correctness, but it is of the order of 5×10^{-13} cm. To obtain some idea of the degree to which these things are remote from our own scale of human dimensions, suppose protons of this "size" to be magnified to the size of a walnut. On such a scale the atom of hydrogen would be several hundred feet in diameter and the nuclear forces between the protons would not be exhibited until they were within three or four inches of each other, the inverse-square law of electrostatic repulsion being perfectly obeyed for all greater distances of separation. On this same scale of magnification the head of a pin would be approximately five million miles in diameter.

Atomic Characteristics of the Primary Physical Forces

The three primary forces exhibited by matter are diagrammatically shown by figure 4. Gravitational forces are very small per unit-quantity of matter, being of importance only when large amounts of matter interact with each other, as in astronomy. Even in engineering mechanics the gravitational attraction between the different parts of a building is neglected. As far as known, or pending an alteration in the meaning of the words "the force of gravity," the Newtonian forces are of no importance in atomic mechanics, being negligibly small in comparison with other atomic forces.

GRAVITATIONAL FORCES (NEWTON-EINSTEIN)	ELECTROMAGNETIC FORCES (COULOMB-MAXWELL)	NUCLEAR FORCES Electron-neutrino forces (FERMI- ?)	
Emission of: Gravitational waves	Emission of: Light quanta (electromagnetic waves)	Emission of: β-rays and neutrinos	
Newtonian forces between masses $1/r^2$	Electric and Magnetic forces between charges $1/r^2$	A) "Ordinary" forces between heavy elementary particles (protons and neutrons) $1/r^2$?	No theory yet proposed Difficulty: Inverse square law? No saturation of forces
----INTRODUCTION OF THE EXCHANGE PHENOMENON----			
Extremely small—not important	Chemical exchange forces e^{-r/r_0}	B) Exchange forces between heavy elementary particles (protons and neutrons) e^{-r/r_0}	Heisenberg and many others Difficulty: Too small by 10^{12}

Fig. 4. Classification of the known physical forces which govern the behavior of all matter. As indicated in the text, the forces which govern the atomic nucleus are not yet well understood, although many of their characteristics have been determined by experiment.

The electrical nature of matter. The forces of cohesion, impact, tension, compression, strength of materials, sound, light, and chemical action are expressions of the electromagnetic forces between atoms and parts of atoms. During the past forty years have been gradually developed the refinements of Maxwell's classic laws of electromagnetic interaction necessary in a sufficiently fine-grained analysis to "see" the separate parts of atoms. These refinements or modifications are expressed in the quantum-theory in its modern form, called the wave-mechanics. As the word quantum indicates, one primary change or modification involves the introduction of an element of discreteness, limiting an infinite number of possible states of motion which might occur to a finite series of steady states between transitions. Atomic behavior corresponds to this discrete or discontinuous type of electromagnetic action. In any large-scale problem, however, the quantum-theory reduces to the familiar laws of electromagnetic theory.

The other primary modification of classic electromagnetic theory necessary in dealing with the mechanics of the outer structure of the atom is the hypothesis of the wave-nature of matter, as expressed in the modern form of the quantum-theory called the wave-mechanics. Developed since 1925, this treats all material particles as "wave-packets." Approaching the scale of atomic dimensions the simple picture of the various particles of matter as mathematical points is inadequate; the assumption that they are endowed with certain wave-characteristics, and the development of equations on this basis using the electromagnetic forces, have permitted almost complete success in dealing with the difficulties previously encountered in the electromagnetic theory of the interactions between atoms and in the outer structures of atoms. Forces in a crystal lattice, forces of chemical interaction, conduction of electricity in metals, electron-diffraction, atomic and molecular spectra, ferromagnetism, and similar problems are examples of the success of this modification. These basic ideas of wave-character and quantization or discontinuity, incorporated in the wave-mechanics, have also proved successful to at least a considerable extent in treating the mechanics of the nuclear forces, which are not necessarily electromagnetic and in fact may be considered at present as intrinsically non-electrical.

Exchange-forces. The phenomenon of exchange refers to the necessity of considering certain effects, which arise since atomic parts are similar. The importance of this phenomenon arises because, according to the wave-mechanics, atomic parts are not just mathematical points but extend somewhat diffusely through a region of space which has a size greater than zero, and hence may overlap each other.

The forces between two atoms, even though considered intrinsically as electrical forces, depend not only on their distances apart, as based on the inverse-square law, but also on the phenomenon of exchange of charges. The latter gives rise to an additional energy-term in the equation, corresponding to the particles exchanging their charges, when they are near to each other, with a certain frequency. This type of exchange is responsible for the so-called chemical forces, which are interactions between the outer parts of atoms causing them to associate together and to form chemical compounds. Furthermore, all atomic particles are endowed with intrinsic spin, or angular momentum, and the exchange of spin can give rise also to a similar additional force. These extra forces are called *exchange-forces*, and they may be attractions or repulsions, depending on specific conditions.

In this connection, there is one fundamentally important requirement which operates to give large effects dependent upon exchange—the Pauli principle. This states that if two particles in an atomic system not only are similar but are indistinguishable or identical in all their properties, they cannot both be in the same "quantum-state," that is, they cannot have the same position, velocity, and orientation. If more than one electron, for example, takes part in exchange, this law exercises a regulatory function with regard to the states which the several electrons can occupy, and thus may determine whether attraction or repulsion will arise. When particles are thus identical, they may exchange their places and properties without actually changing the system under discussion in any known way.

Forces inside atomic nuclei. These exchange-forces are thought of as simply additional attributes of electromagnetic forces which arise when the analysis is made fine-grained enough to deal with atomic dimensions. The forces inside the nucleus might also be arbitrarily considered as further large modifications of the ordinary electromagnetic forces which apply to these still smaller regions of space inside the nucleus. They might equally well be considered as enormous modifications of the law of gravitation setting in abruptly at very short distances between massive particles. The nuclear forces are so tremendously large and set in so abruptly at very short distances that it seems less arbitrary to regard them simply as another kind of force, obeying, however, the same requirements of fundamental discreteness and exchange as were found necessary in the quantum-theory or wave-mechanics obeyed by the electrically charged particles forming the outer structures of atoms. In addition, since some of the nuclear particles possess electrical charge and magnetic moment, the electromagnetic requirements of the wave-mechanics are imposed on the nucleus, for example, as regards the radiation of electromagnetic waves (gamma rays).

"Understanding" the nuclear forces. Figure 4 also indicates the status of the theoretical understanding of the nuclear forces. The nature of these forces is intimately associated with the phenomenon of the emission of beta rays, which are electrons (positive or negative) shot out of radioactive nuclei with a continuous distribution of energies. Fermi has attempted to explain the forces between the heavy particles in the nucleus as arising from exchange-interactions of the electrons and neutrinos associated with the heavy particles. Using the data available, however, these forces appear to be 10^{12} times too small to account for the known energies of binding of nuclei. The shape of the energy-disitrbution of the beta rays predicted by this theory furthermore does not agree with the observed distribution. Numerous modifications of the exchange-phenomenon have been considered, following the example of Heisenberg, involving exchange of charge, exchange of spin, exchange of both charge and spin, and even the exchange of pairs of charges, but none of these appears satisfactory. At present, while there is no weighty experimental evidence which shows that the forces inside the nucleus are specifically of the exchange-type or even largely so, they are generally believed to be of this type, chiefly because forces of an "ordinary" type (derivable from a potential) do not exhibit the property of saturation. This refers to the fact that each particle added to a nucleus is bound to the rest by a total energy which is about the same, per particle, throughout the atomic table. With ordinary forces each particle interacts individually with each other particle already present, and the energy of binding for adding another particle should be much larger if the initial nucleus contains many particles than if it contains only a few. Consequently, to retain the idea of ordinary potential forces, it is necessary to assume forces which do not act independently between pairs of particles, but which are determined by the number of particles present. In this case the interaction between a pair of particles is itself altered when a third particle is added to the system.

The experiments at the Department on proton-proton scattering exhibit a definite deviation in the variation with angle which is roughly in agreement with the assumption of exchange-forces. The deviation of about 10 per cent excess scattering at an angle of 20° from the curves which would be expected on the basis of the wave-mechanics for the simplest kind of scattering (spherically symmetrical) has the sign (indicating a slight added repulsion) which is expected if exchange-forces are acting. The major features of the proton-scattering measurements, however, and of neutron-proton scattering measurements as well, are just as would be expected if an ordinary potential-well exists, giving rise to a strong attraction. Except for differences of detail, they do not indicate whether or not this potential-well represents exchange-forces instead of an ordinary potential. Exact and comprehensive proton-scattering measurements over the range from 200,000 to 5,000,000 volts, or the future discovery of characteristic proton- or neutron-scattering anomalies at higher voltages, may be expected to remove these present uncertainties as to the exchange-nature of the nuclear forces.

Intimately a part of this problem of the nuclear forces is the apparent lack of conservation of energy in the emission of continuous beta-ray spectra by radioactive nuclei (those which spontaneously change from one chemical element to another by the emission of positive or negative electrons). The initial and final nuclei presumably being the same whether a fast or a slow electron is ejected, one must account for the apparent disappearance of some energy when the beta ray is a slow one. Physicists have "solved" this problem by inventing a new particle of mass equal or nearly equal to zero, the "neutrino," whose chief property is the ability to carry off this energy without detection—almost, in fact, without possibility of detection, since it has no other definite property except angular momentum or "spin" (without magnetic moment).

Another aspect of this general problem is the detailed structure of a nucleus. As Bohr has emphasized, this aspect may never be adequately treated, since the sizes of nuclei are not much greater than the "sizes" of the particles of which they are composed. One further difficulty is the present lack of a relativistic equation for the proton or neutron, something like the Dirac equation for electrons. Lack of a proper relativity-correction seriously impedes any efforts toward a sound interpretation of the minor quantitative details of the proton-proton scattering measurements. The interaction of the primary constituents of matter at still higher energies is also a logical extension of the problems encountered in the nucleus. In the region of cosmic-ray energies, certain facts about electrons and radiation appear to be established but otherwise there is a large degree of uncertainty.

It is thus clear that the difficulties in understanding the forces which govern the atomic nucleus, now that much information about those forces has been acquired, constitute a fundamental problem in physics due to a failure of concepts. The most hopeful direction toward a solution of these difficulties seems to be in the extension of quantitative experimental data on two-body heavy-particle interactions (proton-proton and proton-neutron), especially through higher voltages, and the accumulation of similar data on the beta rays.

The Atomic Physics Observatory

This general discussion of the interest and work at the Department of Terrestrial Magnetism in the field of nuclear physics can hardly be concluded without making reference to the new high-voltage equipment the Institution has recently provided the Department for the extension of these investigations.

The new installation, called the "Atomic Physics Observatory," is shown in plate 1. The equipment (plate 2) comprises a steel tank 55 feet high and 37.5 feet in diameter containing an electrostatic generator and vertical high-voltage tube, insulated for constant and accurately controllable potentials to 5,000,000 volts (and perhaps higher) by dry air at pressures to 50 pounds per square inch. The electrical power of the generator is of the order of ten kilowatts. Beneath the tank is an underground target-room which provides earth shielding for personnel and instruments from the powerful intensities of gamma rays and neutrons and is connected with the present high-voltage laboratory by a tunnel.

Perhaps the most difficult technical problem encountered in the design of this equipment has been that of providing a suitable insulating support for the high-voltage electrode. This is a 5-ton steel "ball," 19 feet in diameter. It contains belts and other vibrating machinery and is supported by the outside tank, which might sway slightly during very high winds. This ball was designed to be supported 26 feet above the base-platform on fireproof insulating

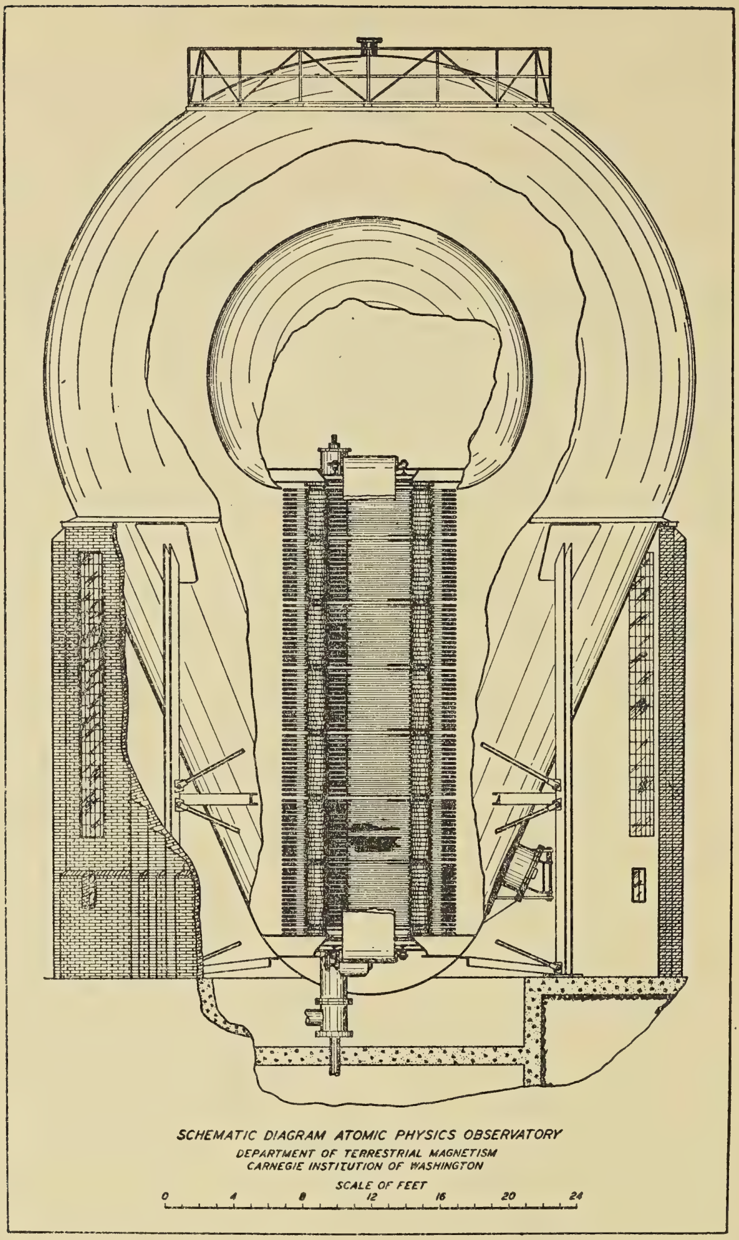
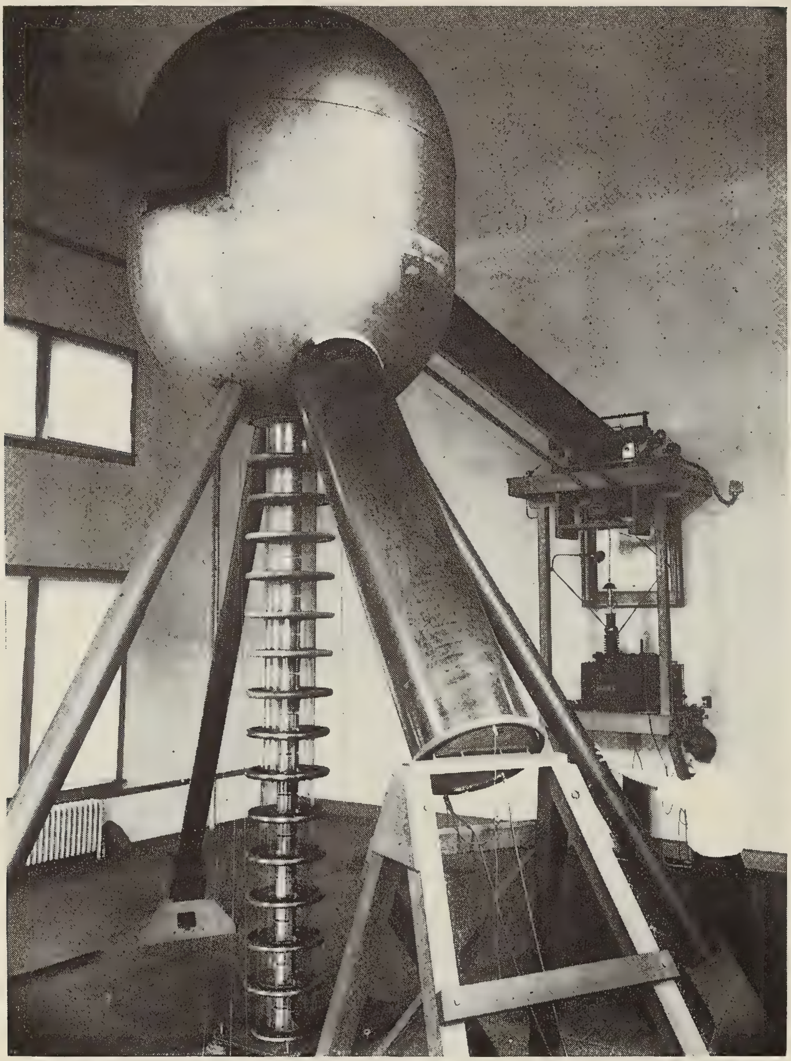

SCHEMATIC DIAGRAM ATOMIC PHYSICS OBSERVATORY
DEPARTMENT OF TERRESTRIAL MAGNETISM
CARNEGIE INSTITUTION OF WASHINGTON
SCALE OF FEET
0
4
8
12
16
20
24

Fig. 5. Schematic diagram showing the internal construction of the new 5,000,000-volt equipment. The targets and observing instruments at the bottom are omitted. The steel ball shown inside the tank can be charged to any desired potential up to 5,000,000 volts. Protons, electrons, or other bombarding particles released at the top of a large vacuum-tube reaching into this ball are accelerated by this potential downward to targets in a room below the structure shown. The high-voltage ball is 19 feet in diameter and is supported on porcelain columns 26 feet high.

columns of porcelain. Porcelain has notoriously bad mechanical properties in bending or tension. To overcome these the adopted design is like a stack of seven flat steel table-tops, each spaced from the next higher by four stubby legs of porcelain; the porcelain in this way is never subjected to any stresses except those of compression. Figure 5 shows a schematic sketch of the internal equipment.

View of the completed Atomic Physics Observatory at the Department of Terrestrial Magnetism of the Carnegie Institution in Washington



The 1,200,000-volt constant-potential equipment at the Department of Terrestrial Magnetism of the Carnegie Institution in Washington. This equipment has been used during the past several years for studies of nuclear transmutations and for direct measurements of the large forces which bind together the component parts of the nuclei of atoms. The targets and observing instruments are in a separate room beneath the high-voltage equipment shown here.

UNIVERSAL ASPECTS OF ATMOSPHERIC ELECTRICITY

O. H. GISH

Assistant Director, Department of Terrestrial Magnetism
The Carnegie Institution of Washington

ELECTRIC forces are at play in the atmosphere all the time everywhere on the Earth. But these forces are not conspicuously manifested except on comparatively rare occasions. The electric state, so prominently heralded by thunder and lightning, is not universal. In the polar regions, a thunderstorm is an event which may be witnessed only a few times in a lifetime. Although thunder may be heard as far as 12 to 14 miles from the source when conditions are favorable and perhaps at a distance of 6 to 8 miles under average conditions, or in an area of 100 to 200 square miles, yet on one-fourth of the Earth's surface thunder is heard on less than 1 per cent of the days, while on about one-eighth of that surface thunder may be heard on 10 per cent or more of all days. The greatest frequency is reported for the region of Java, where 60 per cent of the days may be thundery (1; see "Literature Cited" at end of paper). The average thunderstorm continues, at a given place, for little more than an hour. The electrical manifestations of the thunderstorm, so noticeable without the aid of instruments, are therefore comparatively rare and transitory.

In fair weather, electric forces in the atmosphere can be readily detected with instruments. This fair-weather electric state was first observed by Lemonnier in France (1752), soon after the investigation of the electricity of storms was started under the stimulus of Franklin. Others from time to time verified this observation at various places on the Earth in the following years. The electric force in fair weather was everywhere found to be so directed as to drive positive electricity toward the Earth. Thus this early began to appear to be a universal phenomenon. It is the purpose to present here an integrated view of this and related electrical phenomena of the atmosphere.

Lord Kelvin, who took a keen interest in the subject, urged that "incessant recording" and a sort of "electrogeodesy" are required to solve the riddle presented by the phenomena of atmospheric electricity. He developed instruments and methods for making measurements of the intensity of the electric force so that those made by different observers at various places would be expressed in a common measure. He was the first to inaugurate the continuous registration of the electric potential of the air at Kew Observatory near London, and, using a portable electrometer designed for the purpose, he directly contributed to the "electrogeodesy" by making such measurements at a number of places during his travels. But it was clearly seen by Kelvin and others, including Rowland in America, that the task is one which requires extensive cooperative endeavor. Kelvin and Rowland sought to effect such cooperation by introducing a resolution, which was adopted at the Second International Electrical Congress at Paris in 1884, in which the governments of all countries were urged to make observations of atmospheric electricity and of the electric currents in the Earth. There was some apparent response in a few countries.

Interest in this subject was inactive in America during more than a century following Franklin's pioneer investigations. Beginning in 1861, a frontier German physician, A. Wislizenus, made measurements thrice daily for twelve years at St. Louis, Missouri. Shortly after the Weather Service, of the United States Signal Corps, was inaugurated, several prominent American scholars urged that quantitative measurements of atmospheric electricity be included in the program of that service. In 1879 the American Medical Association appointed a committee and authorized expenditures for the maintenance of atmospheric-electric observation at several stations, in order to investigate the possible bearing of the electrical state of the atmosphere upon human health. This committee proposed to the Chief Signal Officer to stand the expense for the instruments if the Weather Service would undertake the observations (2). Cleveland Abbe, who had served as meteorologist since the establishment of the Weather Service and to whom such matters were referred by General Hazen, then chief signal officer, expressed the view that in addition to such observations being of use for investigating the bearing on human health, it was worth while for the Weather Service to inquire whether they may be an aid in weather-forecasting, especially in the case of thunderstorms and tornadoes.

Abbe was authorized to consult with Professor H. A. Rowland, of the Johns Hopkins University, with a view to developing a plan by means of which observations of real value might be made at stations of the Weather Service. Rowland took an active interest in the matter and, at the first International Congress of Electricians at Paris (1881), moved that the Congress form a commission for the purpose of organizing and directing an investigation of atmospheric electricity. While in Europe, he examined instruments developed there for recording the potential of the air. That developed by Lord Kelvin was then in operation at the Kew Observatory. Rowland, in reporting to Abbe his observations which he made in Europe, expressed the opinion that the atmospheric-electric observations to be effective should be extended to all parts of the world, but indicated his willingness to establish an atmospheric-electric observatory at Baltimore. The subject was also discussed at the Conference of Electricians held in Philadelphia, September 1884, and Rowland, president of the Conference, appointed a committee of prominent American physicists and engineers to consider the subject. Lord Kelvin and Lord Rayleigh, who happened to be in Baltimore, and Rowland were invited to participate in the second meeting of the committee. This committee strongly urged the United States government "to take steps to secure on a larger scale observations of atmospheric electricity and strength of earth-currents."

The Chief Signal Officer had already decided to undertake such work and arranged that some members of the class of observers to graduate from the training course at Fort Myer in June 1882 should be selected for training as special electrical observers. In the latter part of 1884 T. C. Mendenhall was appointed chief of the physical laboratory and instrument-division. Along with his other duties he directed the atmospheric-electric work, even after resigning in 1886 to become president of the Rose Polytechnic Institute, until the work was abruptly discontinued in August 1888.

Photographic registration of the potential at a point in the air was begun at Baltimore on January 1, 1885, and continued until the middle of July 1887, when the physical laboratory was moved to a new building. It was resumed again in the summer of 1888, but was apparently discontinued when the entire electrical program of the Signal Office was abandoned in August of that year. At six other places—Ithaca, New York, Boston, Massachusetts, New Haven, Connecticut, Columbus, Ohio, Terre Haute, Indiana, Washington, D. C.—measurements were made manually on alternate hours, there being five such observations each day, namely, from 9 a.m. to 5 p.m., during the last year. These observations were, however, supplemented at some of the stations by observations made at shorter intervals on days when unusual electric conditions seemed to prevail.

Although the data obtained on this program have some value for the investigation of the general problems of atmospheric electricity, yet the effort spent was not as effective as it might have been without the very unfortunate restrictions which the wording of the appropriation act seemed to place upon the scope of the work. At the outstart it was decided that only such investigations were allowable as would promise an answer to the question, "In the present state of meteorological science can observations of atmospheric electricity be utilized in forecasting the weather?" When a negative answer to this one question was indicated the work was discontinued. Following this the subject was neglected in the United States for nearly two decades longer, but in Europe there was an awakened interest at about this time, which ushered in a new epoch in the history of this subject at about the beginning of the twentieth century.

In January 1902 L. A. Bauer, also realizing, with Kelvin and others, the need for observations on a worldwide scale, submitted to the Carnegie Institution of Washington a "plan for a Bureau of International Magnetic Research," which was later established as the Department of Terrestrial Magnetism of the Carnegie Institution of Washington. The avowed purpose of this organization was "to investigate such problems, of worldwide interest, as relate to the magnetic and electric condition of the Earth and its atmosphere . . .," which was to include as an important aspect "an exhaustive study of the manifold variations of the Earth's magnetism and electricity" (3). With improved methods and equipment for measuring the electrical condition of the atmosphere on the open sea, developed largely under the supervision of W. F. G. Swann, a much more comprehensive program of such measurements than had theretofore been undertaken was started in 1915, at the beginning of Cruise IV of the yacht *Carnegie*. Such observations were carried out on all following cruises with equipment that was improved from time to time, continuous registration of the potential and of the conductivity of the air being successfully employed on the last cruise, which came to a tragic end on November 29, 1929, at Apia, Samoa. These observations were quite well distributed over all oceans from 60° north to 60° south latitude, and constitute the most outstanding contribution to electrogeodesy.

Equipment for continuously registering both the potential and the conductivity of air was developed and tried out at the new laboratory of the Department of Terrestrial Magnetism and, after the principal developmental stage was passed through, registration of air-potential was begun at Washington, D. C., in February 1917 and of air-conductivity in January 1919, and both continued without important interruptions until the latter part of 1932. In October 1929 a more complete program of registrations was started at the Tucson Magnetic Observatory of the United States Coast and Geodetic Survey under a cooperative arrangement in which instruments and some technical supervision were supplied by the Department of Terrestrial Magnetism. These are still being continued.

In the past decade several individuals have helped to extend the atmospheric-electric survey in the United States by making somewhat extended observations of one or more phases of atmospheric electricity. Professor Joseph G. Brown (4), at Stanford University, registered the potential and the net electric charge of the air, beginning in 1929. The potential of the air was registered somewhat intermittently from September 1931 to April 1932 at the University of Colorado, Boulder, Colorado, under the direction of Professor James W. Broxon (5). Registrations or observations with eye-reading instruments, generally for much shorter periods, have been made by several others, namely, C. H. Dwight at Cincinnati University and at Martha's Vineyard (6), J. E. Cooper at Yale University (7), and members of the Department of Terrestrial Magnetism at Lakin, Kansas, 1918 (8), Greenport, Long Island, 1925 (9), Point Loma, California, 1923 (10), Penalosa, Kansas, 1929 (11), Rapid City, South Dakota, 1935 (12), and at various points in the general vicinity of Washington, D. C. (13). Such is the rather tardy and all too meager electrogeodesy of the United States proper, although in Alaska a large program of registrations and manual observations was carried out by the Department of Terrestrial

Magnetism at College, near Fairbanks, in cooperation with the United States Coast and Geodetic Survey and the University of Alaska during the Second International Polar Year, October 1932 to September 1933 (14).

Observations in Europe are more numerous than in other parts of the world. The British Meteorological Office maintains continuous registration of the electrical potential at three places, namely, Kew Observatory—the longest continuous series of observations; Eskdalemuir, Scotland; Lerwick, Shetland Islands. In each of the countries Ireland, France, Spain, Switzerland, Austria, Germany, and Sweden continuous registration of air-potential and some other observations have been made for considerable periods at one or more places. Similar registrations have also been made in various other parts of the world—Egypt, South Africa, India, China, Japan, Brazil, the Argentine, New Zealand, Australia—and the Department of Terrestrial Magnetism has maintained an unusually complete program of registration near Watheroo (Western Australia) since 1922 and near Huancayo (between ranges of the Andes Mountains in Peru) since 1924, and through cooperation with the Department of Scientific and Industrial Research of New Zealand has continued the registration of air-potential at Apia, Samoa. The electrical condition of the air in the polar regions has been investigated on a number of expeditions, the Carnegie Institution cooperating to that end, at least to the extent of providing instruments and training observers for four expeditions.

The atmospheric-electric survey of the world, accomplished in this way, while not yet adequate for an exhaustive study of all the interesting electrical phenomena of the atmosphere, does show beyond reasonable doubt (*a*) that over all fair-weather areas there exists an electric field which is so directed that positive charges in the atmosphere drift toward the Earth while negative charges drift away from it, (*b*) that, although the intensity of the electric field varies from place to place and from time to time, yet in the long run there is no trend either toward an increase or a decrease of the general field, and (*c*) that over the open oceans, where conditions are less complicated and doubtless more representative than at many places on land, the potential of the air 1 m above the surface averages about 130 volts greater than that of the Earth.

When it was learned that the air generally conducts electricity, the universal electric field of fair weather presented a further challenge. It was at once evident that the electric charge of the Earth in these areas would soon be neutralized were it not replenished in some way. How is this charge replenished? Although that appeared as an important question about the beginning of the present century, even today it cannot be answered with assurance.

It is convenient to speak of a supply-current, which in an unknown manner replenishes the negative charge of the Earth by compensating both the positive charge which drifts to Earth in fair-weather areas and the negative charge which drifts towards the high atmosphere. It may be that the mechanism which generates this supply-current is located in areas of storm, and that the supply-current, in the sense here employed, is the net electrical interchange between air and Earth effected in areas of storm by a variety of modes of transfer of electricity.

Because of the complex of factors at play in areas of storm and the impulsive nature of some electrical aspects of storms, it has not yet been feasible to make measurements which are adequate to ascertain definitely whether storms are generators of the supply-current. However, according to this view, the total air-Earth current for all the fair-weather areas of the Earth should be equal to the total supply-current from all storms.

It is now feasible to obtain from automatic registrations the density of the current which is impelled from air to Earth by the electric field in fair-weather areas. This may be done either by a "direct" method in which the current is "collected" on an insulated plate set flush with the Earth's surface or by indirect means which involve the measurement of three factors, namely, (*a*) potential-gradient or field-strength (E), (*b*) the conductivity contributed by the positive ions (λ_1), and (*c*) that contributed by the negative ions (λ_2), the current-density (i) being given by the relation $i = (\lambda_1 + \lambda_2)E$. The sum $(\lambda_1 + \lambda_2)$ shall be called total conductivity or simply conductivity when that entails no ambiguity. The technique for obtaining automatic registration of these factors is now more satisfactory than that for registration by the direct method. Because of this and the advantage for analysis of knowing how the several factors vary, the latter method has been employed in most long series of registration.

The conduction of electricity in air and other gases became a concrete conception in the last years of the nineteenth century. Although Coulomb found in 1785 that an electrically charged body when exposed in air loses charge at a rate given by the law which bears his name, yet this received little attention until Linss in 1887 made measurements three times each day for two years of the proportional rate of dissipation, or the coefficient of dissipation (a), of electricity from a charged body exposed in the open air. The coefficient a is defined by Coulomb's law, namely $dQ/dt = -aQ$ where Q represents the quantity of electricity on the body and dQ/dt represents the rate at which charge is lost. These showed that, in present-day terminology, (1) the electrical conductivity of air, which is roughly proportional to a, varies considerably from time to time, (2) that it is greater in summer than in winter, and (3) that during the year the potential-gradient on the average varies inversely as the conductivity. This inverse relationship is frequently found. It implies that the electric conduction-current in fair weather tends to vary less than at least one of the two component factors, namely, potential-gradient and conductivity.

The conductivity of air was earlier thought to arise from the presence of impurities, such as particles of dust, smoke, fog, or water in other forms, until Elster and Geitel, about 1895, from their numerous measurements of electrical dissipation, showed the reverse, namely, that air generally conducts electricity best when pure and relatively dry. These observations were clarified when the conception of gaseous ions was introduced near the end of the nineteenth century.

But how are ions formed in the open air? In a search for an answer to that question, Elster and Geitel discovered that the air over land generally contains radioactive matter, that most of the important constituents of the Earth's crust contain measurable amounts of radioactive matter, and that the former is doubtless derived from the latter. Since it had recently been found that radiations from radioactive substances form ions in the surrounding air, the conductivity of the air over land seemed to be largely accounted for. But Elster and Geitel and C. T. R. Wilson in the first years of this century found that air from which radioactive matter had been carefully removed continued to be somewhat conductive. These observations apparently stimulated a series of investigations by other physicists, which eventually led to the discovery of what are now commonly called cosmic rays. That this ionizing radiation is of extra-terrestrial origin was first clearly indicated by the observations made by Hess (1911) during ten balloon flights. These observations were verified and extended by Kolhörster in 1913. Many measurements of the cosmic radiation were made on all oceans during cruises of the *Carnegie*, and the intensity was found to be about the same as at sea-level on land (15). Furthermore, many measurements over the open seas showed that the radioactive content of air there is not more than 2 per cent of that found on the average over land (16), and measurements of the amount of radium in sea-water indicated that to be less than 1 per cent of the amount found in the soil (17). These observations strongly corroborated those of Hess and Kolhörster and showed that this radiation is doubtless of universal distribution and constitutes the preponderant ionizing agent over the oceans. Observations made on balloon flights, of the several factors involved, show furthermore that everywhere in the air at altitudes greater than 1 or 2 km, and up to an altitude of at least 20 km, the air is ionized almost exclusively by the cosmic radiation. This radiation is, therefore, an all-important factor in determining the character of the universal aspects of atmospheric electricity. Without the conductivity of the higher atmosphere, which depends upon the cosmic radiation, the electrical phenomena of the atmosphere would doubtless be very different and lightning would probably be more intense and more frequent.

At sea-level the cosmic radiation forms ions in pairs, one positively the other negatively charged, at a rate of 1.5 to 2.0 ion-pairs in each cubic centimeter each second. This is practically the complete rate of ion-formation over much of the ocean-areas and doubtless also over land in the polar regions, but over most land-areas the birth-rate of ions is several-fold greater on account of the additional ionization there by radioactive matter. However, the electrical conductivity of the air over land generally does not exceed that at sea and at some places, especially near large cities, it is much less despite the greater rate of production of ions. In the outskirts of Washington it is about one-seventh the value at sea. This apparent paradox was resolved when it was found that the normal small ions in air which contains certain impurities are transformed into large ions which drift more slowly in the electric field and thus contribute less to the conductivity. These large ions are indeed so very sluggish that if all the small ions were transformed in this way the conductivity would be reduced to a small fraction of its value when the ions are all of the small type.

The quantitative study of the different species of ions in the atmosphere has received much attention in recent years, especially from E. Schweidler and V. F. Hess and associates in Austria, Israel-Kohler in Germany, J. A. McClelland and J. J. Nolan and associates in Ireland, F. J. W. Whipple and associates in England, R. A. Hogg in Australia, and G. R. Wait and O. W. Torreson at the Department of Terrestrial Magnetism in the United States. Relations between the concentrations of ions of the several species and the rate of ion-formation, destruction, and transformation, have been at least roughly established. These reconcile, in the main, the observations that the conductivity is generally reduced when the air is polluted. The entities (Aitken nuclei) which serve in the formation of large ions from small ions are generally more plentiful over land than on the open oceans, but even at sea an average of 800 per cubic centimeter was measured on Cruise VII of the *Carnegie.* These Aitken nuclei appear to constitute a widely distributed factor in atmospheric electricity. The effect of the relatively small concentration of nuclei at sea in lowering the conductivity of air is however not negligible; it certainly amounts to as much as 50 per cent. This effect is, however, not only much less but more constant at sea, in respect both to time and to position, than is the case on land. Because of this circumstance and the additional fact that the intensity of the ion-forming agency at sea, namely, the cosmic radiation, is relatively invariable with time and position, atmospheric-electric conditions at sea are much less complicated than on land and reflect universal aspects more directly.

From numerous measurements of the electrical conductivity and the potential-gradient made at sea during cruises of the *Carnegie*, the average density of electric current from air to Earth in areas of fair weather is found to be 3.7×10^{-16} ampere per square centimeter, which corresponds to one ampere in 110,000 square miles. For the whole area of the Earth this would amount to 1800 amperes. This is thought to be the best estimate that can be made at the present time of the total current which inconspicuously flows from air to Earth in all areas where fair weather prevails. The elusive supply-current is, of course, of the same magnitude.

In searching for the origin of the supply-current, it is desirable to know about how much power is required to maintain this current, or the electric field. This can be calculated if one knows, in addition to the total current (I), the effective electrical resistance (R) or the total voltage (V) between the Earth and the region high in the atmosphere where the electrical conductivity is very great. The power (P) is then derived from the relation $P = I^2R$ or IV. The value of V is indicated by measurements made during balloon flights by European observers.

The renowned scientists Gay-Lussac and Biot are said to have undertaken such measurements in 1804, not long after the first balloon capable of carrying men had been developed. The next attempt to obtain such measurements was made by Lecher, at Vienna in 1885. However, most of the measurements that are of quantitative value were obtained later on 57 flights, 4 of which were made in France, 7 in Austria, and 46 in Germany, after 1890 and prior to the World War (18). In general, the potential-gradient was found to decrease with altitude, rapidly in the first kilometer, then more slowly, until at about 10 km—the highest altitude reached on any of the flights—a gradient of not more than 5 per cent of the value at the surface is indicated. From an approximate integration of the mean observed gradient for the respective altitudes, the potential (V) of the high atmosphere is estimated to be at least 200,000 volts greater than that of the Earth. Some estimates of V exceed this by 50 per cent or more. From the values of I and V estimated in this way it appears that about 400,000 kilowatts of power are required to maintain the universal electric field.

The general decrease of gradient with altitude implies another aspect which is doubtless universal, namely, that the net electric charge of the atmosphere is positive and that the total charge in the lower 10 km is equal to at least 95 per cent of the negative charge on the Earth. This is in turn a necessary result of the increase of air-conductivity with altitude.

The measurement of air-conductivity during balloon-flights was earlier attended with considerable difficulty. That is doubtless the reason why such observations were made on only twelve flights, from 1905 to 1920, eleven of which started in Germany and one in Russia. The maximum altitude at which measurements were made on most of these flights was less than 6 km, but on one flight measurements were made at a height of nearly 9 km.

On the flight of the stratosphere balloon *Explorer II*, sponsored

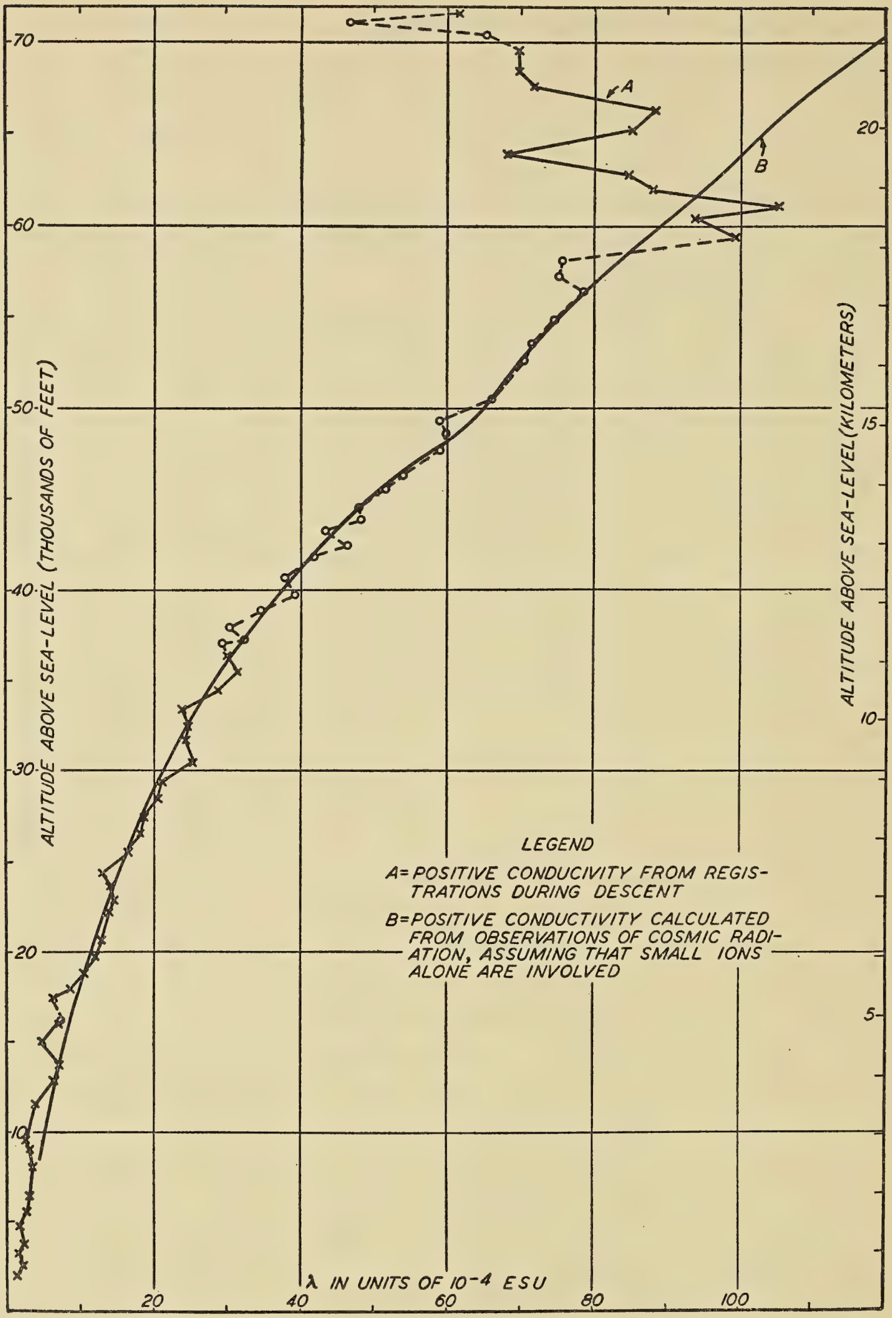


Fig. 1. Air-conductivity, flight of *Explorer II* near Rapid City, South Dakota, November 11, 1935.

by the National Geographic Society and the United States Army Air Corps, the air-conductivity was registered throughout the flight (November 11, 1935) up to an altitude of nearly 22 km, with apparatus designed and constructed at the Department of Terrestrial Magnetism of the Carnegie Institution of Washington (19). On all flights the air-conductivity was in general found to increase with altitude in a manner which may be illustrated by the data obtained on the flight of *Explorer II* and shown graphically in figure 1. Graph *A* is constructed with the altitude as ordinate and the conductivity as abscissa. The conductivity, in general, increases from the surface up to an altitude of 60,000 feet or 18 km, where it is about 100 times the value at the surface. In the altitude-range 18 to 22 km, it varies irregularly but in general decreases with altitude. This last feature was quite unexpected, but may be attributed to the presence there of substances (perhaps Aitken nuclei) which serve for the formation of ions which are more sluggish, that is, attain smaller velocities when subjected to electrical forces, than the small ions which apparently preponderate in the atmosphere below, at least from an altitude of 5 km up to 18 km. This decrease of conductivity is in the same region where ozone was simultaneously found to be especially abundant. Whether this feature is universal, whether it is generally associated with ozone, and whether there are factors in the atmosphere at yet higher levels which effect a similar diminution of the conductivity, are questions which also have important bearing on other than electrical aspects of the atmosphere and which accordingly call for further investigation of the conductivity of the high atmosphere. However, indirect evidence, which will be outlined later, indicates that this diminution of conductivity at high levels either is not universal or else is rather limited in vertical extent. The unexplored region to which this statement applies extends from 22 to about 60 km. Above 60 km the concentration of ions is now being measured at several places on the Earth by quantitative studies of the reflections of radio waves. Such observations demonstrate that the air above that level is extraordinarily conductive.

The resistance of a vertical column of atmosphere of unit cross-section and extending from the Earth's surface up to a certain height is obtained by integrating the reciprocal of the total air-conductivity (the air-resistivity) from the surface up to that height.

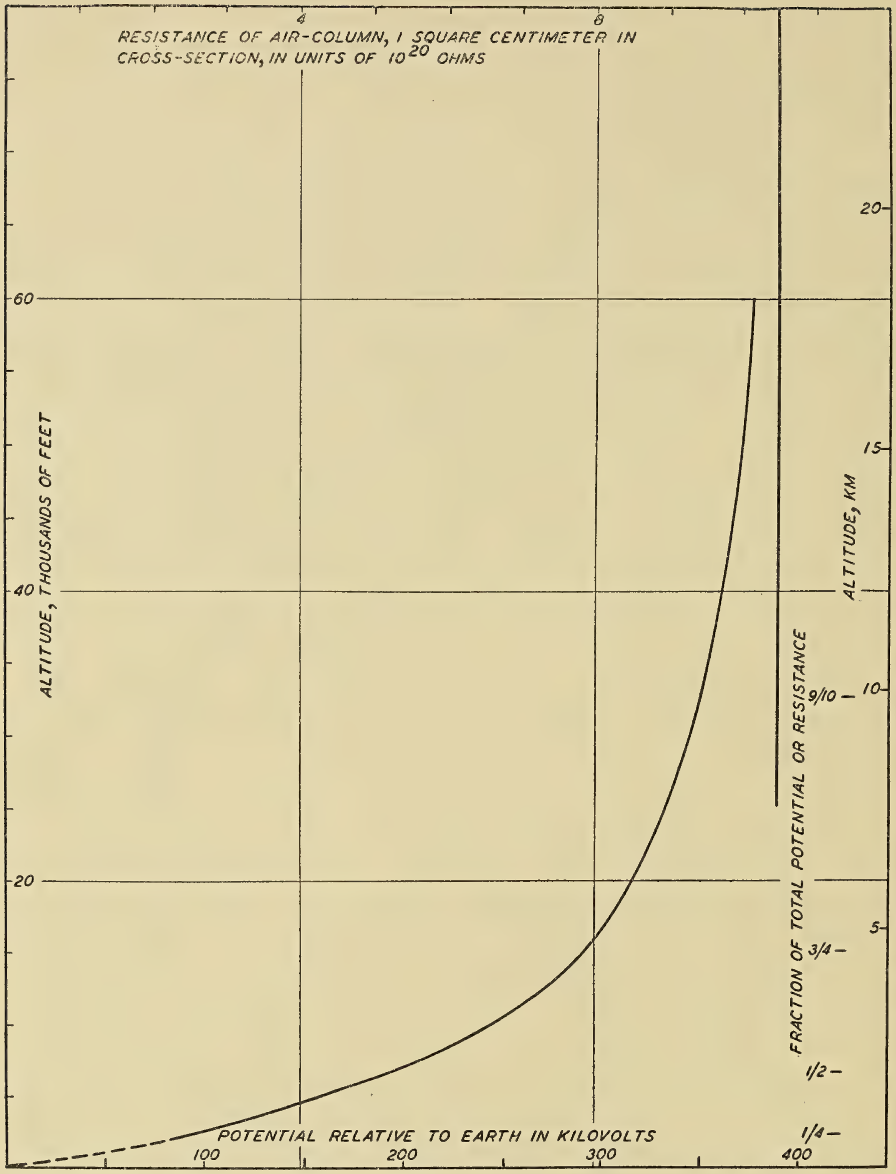


FIG. 2. Potential and resistance between Earth and points in the atmosphere, flight of *Explorer II*, November 11, 1935.

Values of columnar resistance obtained in this way are shown by the graph in figure 2. By referring to the scale at the top of that figure it is to be seen that the resistance of a unit-column extending from sea-level to an altitude of 18 km is about 10^{21} ohms. If this be regarded as representative for the whole Earth, then the total effective resistance from the surface to the 18-km level would be 200 ohms. But most of this is offered by the lower part of the column, the upper 8 km contributing only 5 per cent of the total, whereas the lower 2 km contribute 50 per cent.

Estimates of the columnar resistance up to altitudes considerably greater than 18 km, based on observed values of cosmic-radiation intensity and the indicated trend of that radiation at altitudes not yet explored, indicate that the total columnar resistance up to the ionosphere would not exceed by more than 10 per cent that of a column up to 18 km, provided the small ions at those levels are not transformed into larger, less mobile types, in appreciable number (20). Such estimates may at least set a lower limit to the columnar resistance between the Earth and the ionosphere. An upper limit, not exceeding by as much as 50 per cent the value shown in figure 2 for an altitude of 18 km, seems to be indicated by the extent to which air-pollution of local origin—say from a city like Washington or London—diminishes the air-Earth conduction-current through diminution of air-conductivity in the lower atmospheric strata.

The columnar resistance is apparently 15 to 20 per cent greater near the equator than in middle latitudes (20), owing chiefly to a dependence upon latitude of the vertical distribution of cosmic-radiation intensity. This is interesting because it provides an explanation of the tendency found by S. J. Mauchly (21) of potential-gradient measured at sea to be about 17 per cent less in the equatorial belt than in belts of higher latitude, whereas the air-conductivity at the surface was found to be practically independent of latitude. This latitude-effect in the columnar resistance is not, however, as large as the variation from place to place on land. The columnar resistance over an urban area may be several times that in the relatively pure air over open country.

When all these circumstances are considered it seems likely that the average columnar resistance does not differ much from 10^{21} ohms and that the effective resistance to the vertical conduction-current for the whole Earth is not far from 200 ohms. For this value of R

and for $I = 1800$ amperes, one finds that about 650,000 kilowatts of power is required to maintain the general electric field of the Earth. This estimate is about 80 per cent greater than that which was based on voltage and current, but one is not inclined to attach significance to this disparity. This amount of power is less than one-fourth the rate at which energy is expended in electrical manifestations of a single fairly active thundercloud (22). C. E. P. Brooks (1) from his study of data for thunderstorms of the world estimated that on the average 1800 such storms are in progress on the Earth at any instant. If these estimates are valid, it follows that the power required to maintain the general electric field is only about one-seven-thousandth of that expended in electrical aspects of thunderstorms. Obviously, electrical energy is developed in areas of storm at such a rate that, if only a small part were diverted in such a way as to supply a net electric charge to the Earth and a corresponding positive charge to the high atmosphere, the general electric field could be maintained. The hypothesis that the mechanism which generates the supply-current is located in areas of storm is therefore not ruled out on account of an inadequate source of energy.

Evidence which lends some positive support for this hypothesis is provided by some of the temporal variations in the electrical conditions of fair weather. Although over a long period of years there is, in general, no detectable change, yet both potential-gradient and conductivity vary from year to year, from season to season, and among the changes of shorter duration the regular variation during the day is of especial interest, since it is now established that there is an important universal aspect in this diurnal variation. Although several investigators have expressed the surmise that some other variations in gradient are universal in character, this cannot be taken as an established fact.

In considering the variations from the broadest point of view, it would be desirable to deal with the fair-weather conduction-current, namely, the product of the potential-gradient and the total conductivity. The available data are, however, generally inadequate for this. Over the oceans the average gradient and average conductivity varied significantly from cruise to cruise of the *Carnegie*, but the product of these elements varied to an insignificant extent. Although the average conductivity at sea varies from year to year and from day to day, there is only a relatively small regular variation during the day. The potential-gradient, however, undergoes a pronounced diurnal variation there (23, 24). This is illustrated

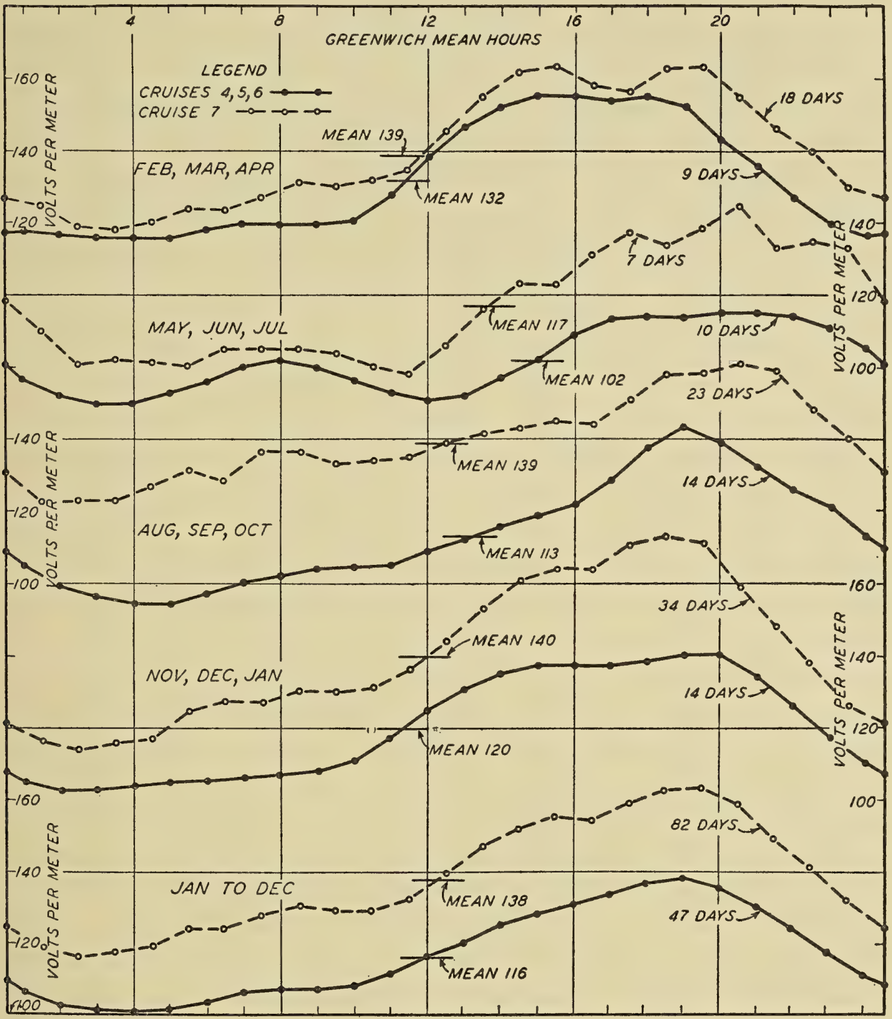


Fig. 3. Mean diurnal variation of potential-gradient from observations on board the *Carnegie*, 1915–1921 and 1928–1929

by the results obtained on cruises of the *Carnegie* which are exhibited in figure 3. Average values of potential-gradient, in volts per meter, are there plotted as ordinates against the hours of the Greenwich day counting from midnight to midnight. The combined results for Cruises IV, V, and VI are shown separately from those for Cruise VII. The latter, although generally the greater, varies in nearly the same manner as the former. Fixing attention on either of the two lower graphs, which represent hourly averages for the year, it is to be seen that the gradient is smallest about 4 hours after Greenwich midnight and greatest at about 19^h. That this feature varies somewhat throughout the year is shown by the other graphs, which represent different quarters of the year. When these same data are averaged for hours of the local day, no significant diurnal variation is disclosed. Thus the change in gradient during the day does not depend upon daylight and darkness or other factors which vary in a manner directly related to the elevation of the Sun at a given place. It follows that the negative charge on the Earth, at least in fair-weather areas at sea, is greatest at about 19^h Greenwich time and least at about 4^h and that the supply-current varies in a similar manner.

This universal variation in potential-gradient is also manifested in the diurnal variation of potential-gradient at most places on land. But factors of local origin there often introduce component variations which may preponderate, especially in or near centers of population where the conductivity of the air is reduced by the polluting substances which are introduced into the air at such places. The concentration of these impurities is, of course, affected not only by the rate at which they are supplied to the atmosphere but by the strength and direction of the wind, by rain, or by other processes which scatter these substances or carry them to or from the place at which the electrical features are considered. Variations of the content of radioactive matter in the air over land also result in variations of the concentration of the small ions, upon which the conductivity of the air chiefly depends. Over land the air-conductivity and the gradient may, therefore, be expected to vary during the day in a manner and to an extent which is often dependent upon a set of local circumstances. The diurnal variation in the electric current from air to Earth is influenced somewhat less by such local circumstances, but there is often a dominant local component in this element. The study of these local aspects of atmospheric electricity and the factors which control them has advanced far enough, in the last decade or so, so that it is beginning to be possible to account for some of them quantitatively, at least in rough measure.

It is therefore with considerable assurance that one now looks upon the Earth as receiving a negative charge, through some unknown mechanism, at an average rate of about 1800 amperes, and upon this supply-current as varying in a fairly regular manner during the day, being least at about 4 hours and greatest at about 19 hours after Greenwich midnight, the range from minimum to maximum amounting to about 35 per cent of the mean—the diurnal variation in general being close to that of the potential-gradient at sea. The extent to which the supply-current varies from day to day, from season to season, and from year to year is, however, not yet determined.

The suggestion that this supply-current is generated in areas of disturbed weather is attributed to C. T. R. Wilson. He thought it likely that negative electricity may be brought to Earth in greater quantity than positive by the several processes known to be active there, namely, by precipitation, by normal electric conduction, by silent discharge—especially from more or less pointed objects, which at times becomes conspicuous as St. Elmo's fire—and by lightning, and that positive electricity is transferred to the high atmosphere by electric conduction, where it is readily distributed about the Earth by virtue of the great conductivity there. The technique and equipment for obtaining adequate direct measures of the net current that results from all these processes at any place on the Earth have not yet been developed. Such measurements as are available have, however, been used to estimate (25) that current. The result is in the right direction but the magnitude is too small. However, the latter certainly cannot be regarded as significant. This estimate is therefore not unfavorable to C. T. R. Wilson's suggestion.

Perhaps the principal support for that suggestion is a representation, presented by F. J. W. Whipple (26), of the variation during the Greenwich day in the aggregate of thunderstorms that are in progress on the Earth. This is shown in figure 4, where the total area in which thunderstorms are estimated to be in progress at a given instant is plotted as ordinate against the time of day counted from Greenwich midnight as abscissa. By comparing this graph with those at the bottom of figure 3 it is to be seen that the correlation is good. The supply-current, as indicated by the potential-gradient at sea, is greatest in those hours when thunderstorms are of most widespread occurrence, and this relation seems to hold quite well for most features of the diurnal variation. It is an interesting corollary that, if the thunderstorm is the seat of the supply-current, then a record of the air-Earth conduction-current, at a place which is relatively free from disturbing local conditions, may also be regarded as a record of the total thunderstorm activity of the Earth and that such records should be of value in a study of world-meteorology.

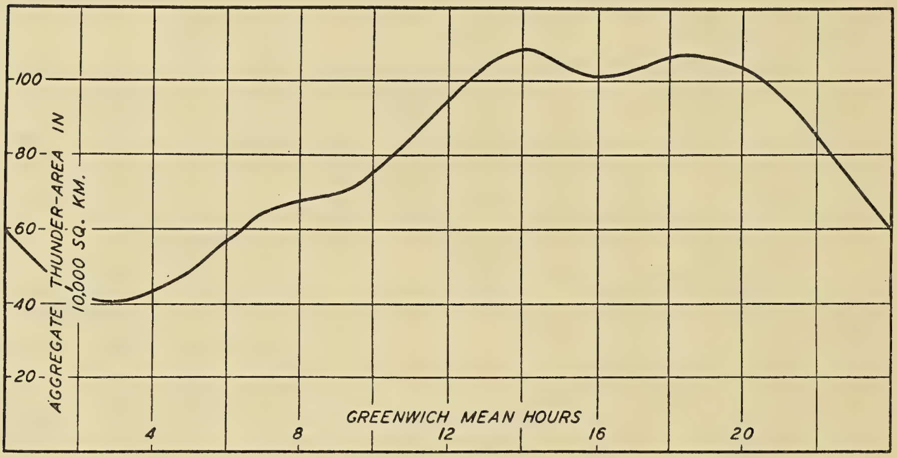


Fig. 4. Diurnal variation of total thunder-areas of the Earth (after Whipple).

The data used by Whipple were assembled by C. E. P. Brooks, who called attention to certain marked deficiencies which, as is readily understood, are to be found in such data. There are vast areas in which very few if any systematic records of the occurrence of thunderstorms have been kept. This applies with especial force to the Southern Hemisphere and to the oceans in general. The extent to which this and some assumptions which underlie Whipple's representation affect the evidence for C. T. R. Wilson's suggestion cannot be decided at the present stage of these investigations. However, there seems to be some supporting evidence here. In any case, unless and until it is proved that the mechanism which provides the supply-current is not located in areas of storm, this view will doubtless stand as the outstanding working hypothesis for the interpretation of universal aspects of the electric field and current in the atmosphere.

Numerous other theories regarding the origin of the supply-current have been proposed from time to time. Apparently every known physical mechanism, and some others, have been invoked. All these have been definitely refuted, except for a few theories in which it has been seriously considered as necessary to postulate departures from classical physics, or some fundamental entities or processes not yet known to physics, in order to account for the universal aspects of atmospheric electricity and terrestrial magnetism (27, 28, 29).

LITERATURE CITED

1. Brooks, C. E. P., *Geophys. Mem.*, No. 24 (1925).
2. Mendenhall, T. C., *Mem. Nat. Acad. Sci.*, vol. 5, pp. 113–318, 1889.
3. *Terr. Mag.*, vol. 9, pp. 1–8, 1904.
4. *Terr. Mag.*, vol. 35, pp. 1–15, 1930; vol. 38, pp. 161–173, 1933.
5. *Terr. Mag.*, vol. 39, pp. 7–22, 1934.
6. *Terr. Mag.*, vol. 38, pp. 309–312, 1933.
7. *Physics*, vol. 7, pp. 387–394, 1936.
8. Bauer, L. A., H. W. Fisk, and S. J. Mauchly, *Terr. Mag.*, vol. 24, pp. 87–97, 1919.
9. Ault, J. P., S. J. Mauchly, and R. H. Goddard, *Terr. Mag.*, vol. 30, pp. 125–146, 1925.
10. Johnston, H. F., *Terr. Mag.*, vol. 29, pp. 13–22, 1924.
11. Wait, G. R., *Terr. Mag.*, vol. 35, pp. 137–144, 1930.
12. Sherman, K. L., and O. H. Gish, *Terr. Mag.*, vol. 42, pp. 289–299, 1937.
13. Johnston, H. F., and G. R. Wait, *Terr. Mag.*, vol. 36, pp. 33–40, 1931.
14. Sherman, K. L., *Terr. Mag.*, vol. 42, pp. 371–390, 1937.
15. *Researches Dept. Terr. Mag.*, Carnegie Inst. Wash. Pub. No. 175, vol. 5, pp. 421–423, 1926.
16. Mauchly, S. J., *Terr. Mag.*, vol. 28, pp. 61–81, 1923.
17. Hewlett, C. W., *Terr. Mag.*, vol. 22, pp. 173–181, 1917.
18. Gish, O. H., *Trans. Amer. Geophys. Union, 17th Annual Meeting*, pp. 182–186, 1936.
19. Gish, O. H. and K. L. Sherman, Nat. Geog. Soc., Contrib. Tech. Papers, *Stratosphere Ser.*, No. 2, pp. 94–116 (1936).
20. Gish, O. H. and K. L. Sherman, *Trans. Amer. Geophys. Union, 19th Annual Meeting*, pp. 193–199, 1938.
21. *Researches Dept. Terr. Mag.*, Carnegie Inst. Wash. Pub. No. 175, vol. 5, pp. 387–424, 1926.
22. Schonland, B. F. J., Atmospheric Electricity, pp. 81–84. London, Methuen, 1932.

23. Mauchly, S. J., *Phys. Rev.*, vol. 18, pp. 161–162, 1921; *Terr. Mag.*, vol. 28, pp. 61–81, 1923.

24. Parkinson, W. C., and O. W. Torreson, *Trans. Stockholm Meeting, 1930, Internat. Union Geod. Geophys.*, Assn. Terr. Mag. Electr., Bull. No. 8, pp. 340–345, 1931.

25. Wormell, T. W., *Proc. Roy. Soc.*, A, vol. 127, pp. 567–590, 1930.

26. Whipple, F. J. W., *Quart. Jour. Roy. Meteorol. Soc.*, vol. 55, pp. 1–17, 1929.

27. Swann, W. F. G., *Phil. Mag.* (London), vol. 3, pp. 1088–1136, 1927.

28. Anderson, W., *Ztschr. f. Physik*, vol. 42, pp. 475–478, 1927; vol. 44, pp. 376–384, 1927.

29. Schweidler, E., Die Aufrechterhaltung der elektrischen Ladung der Erde. Hamburg, Henri Grand, 1932.

THE GENERAL MAGNETIC FIELD OF THE EARTH AND ITS SECULAR VARIATION

J. A. FLEMING

Director, Department of Terrestrial Magnetism
The Carnegie Institution of Washington

The origin of the general magnetic field of the Earth presents the most fundamental and difficult problem of terrestrial magnetism. Despite several centuries of speculation and scientific research there is yet no adequate explanation of how the Earth became magnetic or why it remains so. William Gilbert, in 1600, pictured the Earth as a sphere of magnetic substance; others thought that it owed its magnetization to electric currents flowing beneath its crust because of their own inertia or because of some sustaining agency. Whether either of these views is correct or whether the Earth owes its magnetism to some undiscovered property of matter no search so far has let us know.

Closely associated with the basic problem and rendering its solution more difficult is the fact that the Earth's magnetism is continually changing, slowly when compared with the life span of a human being but very rapidly when compared with astronomical or geological scales of time. Nor are these secular changes small. During the past half-century a gradual change amounting to 10 per cent of the maximum value of the Earth's field has occurred off the Guinea Coast. Were this change to continue unabated the south magnetic pole of the Earth would be in the Middle Atlantic Ocean in less than a thousand years. Clearly, secular change involves the general magnetic field of the Earth and any attempt to explain either phenomenon must also account for the other.

The attack on a problem of such broad scope requires a special type of research and, in a unique degree, world-wide coordination of data and experiments. No single well-planned experiment or observation can supply its solution. Observations must be made in all parts of the world and continued over a lengthy period.

Techniques for the organization and interpretation of these data must be developed and experimental researches must be conducted along lines which will supply information on basic physical properties related to the subject. Recognition of the breadth of the problem caused the Carnegie Institution to establish its Department of Terrestrial Magnetism upon the recommendation of Louis A. Bauer, who became its first director in 1904, "to investigate such problems of world-wide interest as relate to the magnetic and electric conditions of the Earth and its atmosphere, not specifically the subject of inquiry of any one country, but of international concern and benefit."

Acquisition of Data

The first step in the prosecution of this elaborate program was to improve the data upon which knowledge of the Earth's magnetism is based. Even in the days of Gauss sufficient data had been obtained to permit a rough analysis of the field into spherical harmonics. This analysis had led to the important conclusion that, within the limits of reliability of the data, the entire field was due to causes beneath the Earth's surface. With extension of the frontiers of scientific exploration more data became available and the analyses of Gauss were improved by his successors including Adams and Neumayer. In 1885 Ad. Schmidt performed a more precise analysis which confirmed the main conclusion of Gauss but revealed that an appreciable portion of the field was due to causes above the Earth's surface while another portion could be explained by vertical electric currents flowing through the surface of the Earth. But even at that time there were extensive portions of the Earth for which magnetic data were meager or lacking altogether. Were these so-called "external" and "nonpotential" portions of the field real or were they merely the results of employing inaccurate and insufficient data?

To obtain observations of the direction and intensity of the Earth's magnetism over the Pacific Ocean where they were particularly needed, the newly formed Department of Terrestrial Magnetism in 1905 chartered the sailing vessel *Galilee*. Measurements made during three cruises of that vessel, covering 64,000 nautical miles in the Pacific Ocean, revealed such great discrep-

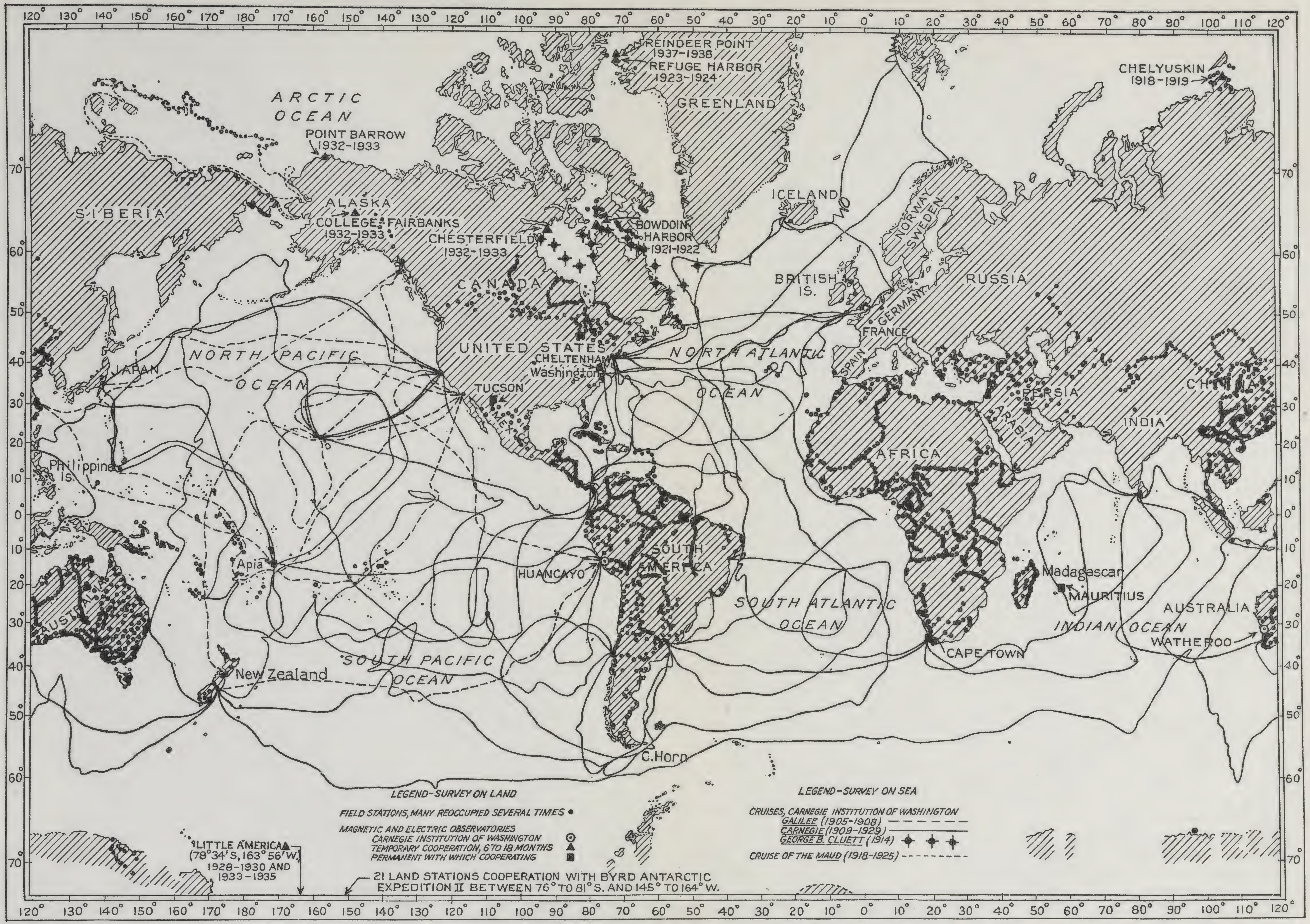

ARCTIC OCEAN
POINT BARROW 1932-1933
REINDEER POINT 1937-1938
REFUGE HARBOR 1923-1924
GREENLAND
CHELYUSKIN 1918-1919
SIBERIA
ALASKA
COLLEGE FAIRBANKS 1932-1933
CHESTERFIELD 1932-1933
BOWDOIN HARBOR 1921-1922
ICELAND
NORWAY
SWEDEN
CANADA
BRITISH IS.
RUSSIA
GERMANY
FRANCE
SPAIN
UNITED STATES
CHELTENHAM
Washington
NORTH PACIFIC OCEAN
NORTH ATLANTIC OCEAN
JAPAN
TUCSON
MEX.
PERSIA
ARABIA
CHINA
INDIA
AFRICA
Philippine Is.
Apia
HUANCAYO
SOUTH AMERICA
SOUTH ATLANTIC OCEAN
Madagascar
MAURITIUS
AUSTRALIA
INDIAN OCEAN
WATHEROO
CAPE TOWN
SOUTH PACIFIC OCEAN
New Zealand
C.Horn
LEGEND-SURVEY ON LAND
FIELD STATIONS, MANY REOCCUPIED SEVERAL TIMES ●
MAGNETIC AND ELECTRIC OBSERVATORIES
CARNEGIE INSTITUTION OF WASHINGTON ⊙
TEMPORARY COOPERATION, 6 TO 18 MONTHS ▲
PERMANENT WITH WHICH COOPERATING ■
LEGEND-SURVEY ON SEA
CRUISES, CARNEGIE INSTITUTION OF WASHINGTON
GALILEE (1905-1908) - - - - - -
CARNEGIE (1909-1929) ———
GEORGE B. CLUETT (1914) ✦ ✦ ✦
CRUISE OF THE MAUD (1918-1925) - - - - - - - -
LITTLE AMERICA ▲ (78° 34′ S, 163° 56′ W.) 1928-1930 AND 1933-1935
21 LAND STATIONS COOPERATION WITH BYRD ANTARCTIC EXPEDITION II BETWEEN 76° TO 81° S. AND 145° TO 164° W.
120° 130° 140° 150° 160° 170° 180° 170° 160° 150° 140° 130° 120° 110° 100° 90° 80° 70° 60° 50° 40° 30° 20° 10° 0° 10° 20° 30° 40° 50° 60° 70° 80° 90° 100° 110° 120°
70° 60° 50° 40° 30° 20° 10° 0° 10° 20° 30° 40° 50° 60° 70°

Fig. 1. World magnetic and electric survey, Department of Terrestrial Magnetism, 1905–1938.

ancies between the actual magnetic elements over that region and the charts used by mariners that construction of a special nonmagnetic vessel, the *Carnegie*, for surveying the oceans was authorized in 1909. Constructed almost entirely of nonmagnetic materials in order that measurements made on board should be as accurate as possible, she stands unique in the history of ships. The *Carnegie* cruised 298,000 nautical miles in all oceans from 80° north to 60° south before her unfortunate destruction by fire in 1929. Not only were magnetic observations included in her program but also measurements of atmospheric electricity, cosmic radiation, meteorology, and, on her seventh and last cruise, investigations of the physics and biology of the oceans.

Although intended primarily for scientific study, the results of the *Carnegie's* observations were of such great practical importance that they were communicated to maritime organizations each time she made port. New charts were constructed for the navigator giving more exact values for the deviation of the compass than were ever before attainable. Today magnetic charts of the high seas are based largely on these observations.

The importance of continuing the *Carnegie's* work is attested to by the recent decision of the British Admiralty to build and commission a similar nonmagnetic vessel, the *Research*, which is now nearly completed. A movement in the United States urging our government to construct a nonmagnetic ship to chart our coastal waters is gaining increasing force. The continuation of magnetic work at sea, while having great utilitarian value, is of even greater importance to the scientific problems of terrestrial magnetism.

To acquire data from the extensive land-areas for which they were needed some 200 expeditions have been undertaken by the Department, beginning in 1905, to various parts of the world, principally to remote and unexplored regions where there were no existing agencies capable of undertaking the work. Many of these expeditions went into deserts, jungles, and arctic regions. Cooperation with many polar expeditions to secure magnetic data was achieved, and assistance was given many agencies in remote portions of the world to enable them to secure the desired magnetic data (see figure 1). This policy of active cooperation is still being maintained.

This extensive program required the development of special instruments. Light-weight magnetometers were designed and constructed for the expeditions on land, which enabled observers to measure with precision the Earth's magnetism in absolute units, referred directly to standards of mass, distance, and time without recourse to intermediate standards. One of the Department's magnetometers, after comparison with the standard instruments of many countries, was established as the international magnetic standard. An accurate sine-galvanometer was designed, constructed, and employed to refer measurements made with the standard magnetometer to the established standards of electric current. By this means magnetic observations in all parts of the world have been rendered strictly comparable. A new electromagnetic instrument, now under construction, will permit measurements of the vertical and horizontal components of the Earth's magnetic field in terms of accepted standards of electric current to an accuracy better than one part in 50,000.

The work of the *Carnegie* and the *Galilee* required development of instruments suitable for use at sea. Investigations of the effect of the motion of the vessel on the readings of the instruments were undertaken. The special marine instruments designed at the Department have been adopted by the British Admiralty for use on the *Research.*

Interpretation of Data

The new data accumulated over these years warranted a new analysis of the Earth's field. Making use of the data obtained by the *Carnegie* and by the expeditions on land of the Department, Furner and Dyson in England and L. A. Bauer in this country independently performed analyses for the epoch 1922. Bauer's analysis, which was performed with the greater refinement, confirmed the main features of Schmidt's analysis for 1885, namely, that the observed magnetic field could be explained only by separation into three portions, the major portion of about 95 per cent due to causes beneath the Earth's surface and the two minor portions totaling about 5 per cent, of which one is due to causes above the Earth's surface and one is such as would be due to vertical electric currents passing through the Earth's crust. Thus the acquisition and interpretation of new data demonstrated that the remarkable results obtained by Schmidt did not arise entirely from paucity of data.

An important result of the new analyses was that they revealed a decrease in excess of 2 per cent since 1885 in that portion of the Earth's field which may be represented by uniform magnetization of the entire Earth and which comprises by far the greatest part of the field. Secular change had been known for a long time, but that the entire magnetization of the Earth should be decreasing at a rate of 1 per cent in 17 years appeared a surprising result. Examination of the data from various portions of the Earth where observations had been repeated, particularly for the interval 1920–1925, substantiated these findings. Taking the average for the entire Earth as far as possible, the intensity of the horizontal component of the field shows a marked decrease during this interval.

The most outstanding feature revealed by repeated observation at the same station is the predominantly local character of secular variation. Extensive areas of the Earth experience a strong shift of the compass to eastward while adjacent areas experience a reverse change; in some few regions the horizontal component of magnetic intensity is increasing while in most regions it is decreasing. Except for the general decrease of horizontal intensity previously noted no systematic change can be detected. Mapping of the isopors, or lines of equal magnetic change, by the Department disclosed a large number of isoporic foci where the change in the magnetic elements is taking place most rapidly. Comparison of the isoporic foci for different periods shows that they are not constant but gradually changing in intensity and configuration, and that their centers slowly shift position. Old foci die out and new foci develop; even the change of the Earth's magnetic field is continually changing. Yet some of the foci have persisted for a very long time; the focus in the region of South Africa, where horizontal intensity is now decreasing at the rate of nearly 1 per cent of its present value each year, has persisted for at least a century. Continuation of this rate of change for another century and a half would cause the compass-needle to reverse its direction in that region.

In order to obtain a more continuous representation of secular change than is supplied by observations at a large number of points reoccupied at intervals of several years, the records produced by permanent observatories have been investigated. Two such observatories have been established and are maintained by the Department, one at Watheroo, Western Australia, and one at Huancayo, Peru (see figure 1). In addition, the Department cooperates in the maintenance of several other observatories. Records obtained by photographic registration at the permanent observatories portray the daily and erratic fluctuations in terrestrial magnetism and also supply correct integrated values of the Earth's magnetism at those places for each hour, day, month, and year. Single observations made at field-stations may be affected by fluctuations taking place at the time of observation, a source of error avoided by continuous photographic registration. Although the permanent observatories are not spaced closely enough to delineate the isoporic foci, the results of their observations substantiate conclusions drawn from field-observations. They also reveal that secular change is a gradual and continuous process. Because of their greater accuracy an 11-year period in secular variation has been detected which is associated with the sunspot-cycle. Recent investigations at the Department have shown that this 11-year period involves that portion of the magnetic field which has its origin above the Earth's surface and that the sunspots are not responsible for the major features of secular change which have their origin beneath the Earth's surface.

The observations of the Earth's field obtained during the past half-century, to which the Department contributed so extensively, have supplied a detailed picture of the present state of the Earth's magnetic field and its prevailing rate of change. They have shown that the Earth may be regarded, to a close approximation, as a uniformly magnetized sphere with its magnetic axis making an angle of 11°.5 with the axis of rotation. Although but feebly magnetized as compared with the magnetization attainable in high-grade magnet steels, the average intensity of magnetization per cubic centimeter for the entire Earth is roughly 10 to 100,000 times as great as the magnetization of ordinary crustal rocks. Such magnetization would be produced by the presence of one magnetized steel embroidery-needle in every five cubic centimeters of the Earth's substance. Appreciable irregularities in the field exist, but they do not cause great departures from the field which would be produced by the hypothetical uniform magnetization.

Secular change could be produced by changing the intensity of magnetization of the Earth's substance beneath certain restricted regions. To return to the embroidery-needle analogy, such a change as has occurred off the Guinea Coast from 1885 to 1922 could be produced by placing 500,000 embroidery needles beneath each square centimeter of the Earth's surface at the center of change with their north-seeking ends pointing downward. Assuming the substance of the Earth to be uniformly magnetized, this change would involve an alteration of magnetization to a depth of over 1000 km.

Attempts to Account for the General Field

The simplest explanation of the Earth's magnetic field was offered in 1600 by William Gilbert, who likened the Earth to a lodestone and defended the view with a model cut from that material. Belief that the interior of the Earth is at a high temperature, derived from the hypothesis of a gaseous origin and from the observation that temperatures in the Earth's crust increase with depth, makes the retention of this simple view difficult since under ordinary conditions all known substances lose their ferromagnetic properties at temperatures which should be encountered at the depth of 15 or 20 km. Unless it can be shown that the high pressures which exist at great depths in the Earth raise the temperature at which substances lose their magnetism or that the Earth's interior is not at the high temperature which is usually assigned to it, the hypothesis of a ferromagnetic Earth must be completely abandoned.

A hypothesis to account for lower temperatures for the Earth's interior has recently been developed in the Department. Suggested by the fact that data from terrestrial magnetism show that the electric conductivity of the Earth at a depth of several hundred kilometers is a hundred times or more greater than near the surface, this hypothesis proposes that the thermal conductivity may also be much greater at great depths. Accordingly cooling of the interior would have proceeded much more rapidly and would be advanced to a greater degree than is indicated from calculations based on extrapolation of observed conditions at the surface. Judgment on such speculative hypotheses must await the acquisition of very definite evidence before they can be either accepted or rejected.

Electric currents flowing at great depths in the Earth have been suggested as a possible cause for the magnetic field. Lamb has shown that appropriate electric currents once started in a sphere of copper the size of the Earth would die out very gradually, falling to one-third their value in about ten million years. But there is no evidence to show that the Earth's conductivity, which determines the rate of decay for the currents, is at all comparable with that of copper. Furthermore, even if the Earth's conductivity were as great as that of copper, the magnitude of these currents at the beginning of geological time must have been preposterously great in accordance with the theory. Several mechanisms have been proposed for sustaining these currents but they are purely speculative, have certain theoretical weaknesses, and at the present stage of scientific attainment are incapable of experimental verification. Furthermore extensive researches conducted by the Department on the natural electric currents which flow in the Earth's crust have failed to give any evidence in favor of the existence of electric currents which would produce the Earth's general magnetic field.

A number of attempts have been made to explain the Earth's magnetism as arising from its rotation. Such attempts have in support of them the fact that the magnetic axis does not differ greatly from the rotational axis. Whether or not the proximity of the magnetic and geographic poles is significant in this connection is still a matter of debate. The fact that they are appreciably separated dooms any purely rotational theory to failure on *a priori* grounds; for even if one of these theories succeeded in explaining the portion of the Earth's field which is symmetric about the axis of rotation it would still leave unexplained the remaining 20 per cent which is perpendicular to the axis of rotation. The various irregularities of the field would also remain unexplained.

However, it is not inconceivable that rotation may have been the original cause of the magnetic field, the magnetic and geographic poles having become separated subsequently. One of the most promising suggestions of this sort and one which is capable of experimental verification is the gyromagnetic theory. This theory supposes that each magnetic particle in the Earth aligns itself with the axis of rotation acting under the same forces as those influencing the behavior of a gyroscope. Results of experimental tests of this theory will be discussed in the next section.

The discovery that a portion of the Earth's field originated from causes above the surface seemed at one time a most difficult fact to explain. The problem does not seem nearly so difficult today because it has been proved that the daily variations and the erratic fluctuations known as magnetic storms also originate above the Earth's surface and are due, according to most evidence, to electric currents in the atmosphere and possibly in outer space. Radio investigations of the electric state of the outer atmosphere, initiated by this Department and other organizations in 1926, have shown that the outer atmosphere is capable of conducting such electric currents as seem required to explain the external magnetic field. (See pages 223–247 for a fuller account of these researches.)

No theory has ever been offered to account for the so-called nonpotential portion of the Earth's field, that is, that which would be produced by vertical currents flowing vertically through the Earth's crust. Partly in the hope of revealing such a current the atmospheric-electric program of the Department was undertaken. Although the presence of vertical currents in the atmosphere has been demonstrated the distribution of these currents does not agree with what is called for by the distribution of the nonpotential field. Furthermore the atmospheric-electric currents are 10,000 times too small to produce the reputed magnetic effects. At the present time most magneticians are inclined to suppose that the nonpotential field does not really exist but appears because of inaccuracies in the magnetic observations. It must not be overlooked that in dealing with large-scale phenomena such as terrestrial magnetism some principles of physics which are of trivial importance in the laboratory may assume measurable importance. Accordingly the possibility that the nonpotential field is a physical reality cannot be completely discarded until its nonexistence is proven.

Attempts to Account for Secular Variation

Any theory of secular variation depends upon the theory of the general magnetic field since the observations show that a large part of what is regarded as the general field partakes of secular variation. In accordance with a ferromagnetic theory secular variation would indicate an alteration of the intensity of magnetization in some of the Earth's substance; electric-current theories would attribute secular variation to alteration of the lines of current-flow either through changes in the electric conductivity of the Earth or in the electromotive forces giving rise to the currents. Theories that ascribe the Earth's magnetism to its rotation admit of no explanation for secular variation.

Changes in temperature for different parts of the Earth have been suggested as a possible cause of secular variation. Such changes would give rise to magnetic changes in accord with either a ferromagnetic or an electric-current theory of the general field. These changes would have to affect a very large amount of material and consequently would have to extend to great depths because of the large amount of magnetic matter required to produce the observed secular change.

Attempts to associate the main features of secular change with observable geological structure or geological changes have not met with success. Secular change progresses with much too great rapidity to suggest a purely geological origin and its magnitude is much too great to have its origin in the relatively thin region which is the principal subject of geological investigation. Present geological regions and centers of geological activity have persisted for a relatively long time, while if the present pattern of secular change had persisted for the past one or two thousand years the Earth's magnetic field at that time would have been much more irregular and heterogeneous than it is now. On the other hand there may be some significance in the fact that the isoporic foci are located principally in the portion of the Earth containing the greatest land-areas—the portion where seismological evidence indicates the presence of a thick granitic layer. Under any circumstances observation of secular change indicates that vast and comparatively rapid changes are taking place at great depths in the Earth which are not revealed by any other geophysical science.

Ad. Schmidt, considering the several analyses by spherical harmonics of the data for different epochs from 1830 by Gauss and Petersen, Adams, Schmidt, Fritsche, Dyson, and Bauer, thinks the simplest and most proximate hypothesis is that the Earth's interior rotates slowly and uniformly against the Earth's crust with a period approximating 500 years. Astronomical and other geophysical phenomena support the idea of displacement of the crust relative to the Earth as a whole. How, under conditions of the interior, as regards pressure and temperature as at present conceived, the iron-nickel core may be specially magnetized is an open question. However, assuming such a rotating core to be a homogeneously magnetized sphere, the periodically changing induced magnetization in the Earth's crust at different places, depending upon its local character, may produce anomalies in the phenomena observable only on the surface. Considering only the directional first terms of the magnetic analyses, the declination and inclination would perform simple harmonic oscillations depending chiefly upon latitude and phase of longitude and upon position of rotating axis of the core. Schmidt finds, to satisfy the few long-period data (London, Paris, and Rome), such a rotation-axis would be directed toward a point in latitude 30° north and in longitude 2° or 182° east—decision on which longitude can only be determined when data for long-period secular variation of intensity become available. The component of the moment of such a core perpendicular to the axis of the rotation, which is alone effective, is 0.19 of the Earth's magnetic moment (M). (According to Bauer's analysis of the Earth's field M is 8.04×10^{25} cgs, the axial component M_p being 7.88×10^{25} cgs, and the equatorial component M_e being 1.60×10^{25} cgs, that is, $0.20\ M$.) Schmidt's discussion and comparison of results based on this hypothesis with the observed values indicate an eccentric position of the axis of such a magnetic core displaced in the direction of Europe. The tenability of such a hypothesis must depend upon empirical knowledge of secular variation, knowledge patently to be had only through further accumulation of data with long-continued series of observations at many well-distributed secular-variation stations.

Utilizing homogeneous material at fourteen observatories, Bartels draws a negative conclusion as regards such a telluric explanation of secular variation. However, his discussion seems not wholly convincing in several particulars and especially so because the interval covered, 1902 to 1920, is so short.

The 11-year period in secular variation which involves the external portion of the magnetic field finds as simple an explanation as the existence of the external field itself. Its close correspondence to the sunspot-cycle arises because of magnetic storms, which are most numerous when sunspots are most numerous. Magnetic storms depress the average value of the horizontal component of the Earth's magnetism so that years near the maximum of sunspots show a lower value for this magnetic element than other years. However, magnetic storms produce no lasting effects on the general magnetic field of the Earth.

Investigations in the Laboratory

Few problems of terrestrial magnetism are suitable for investigation by experiments in the laboratory, but the effectiveness of experimental methods recommends their use whenever possible. Such problems are usually concerned with the basic properties of matter since the large-scale phenomena of terrestrial magnetism cannot be duplicated in experiments.

The gyromagnetic effect was submitted to experimental test at the Department to ascertain if it could account for the general magnetic field of the Earth. Ferromagnetic substances were rotated at high velocity and accurate measurements made of any magnetic change that took place. The effects predicted by theory were verified by experiment: the bodies did become magnetic by rotation, but, in further confirmation of the theory, the effects were too small to be of any importance to terrestrial magnetism.

Possibility that increases in pressure might raise the temperature at which substances lose their ferromagnetic properties was also submitted to test in the laboratory in a joint project of the Department and the Geophysical Laboratory of the Institution. Various substances, including ordinary iron, meteoric iron, and magnetite, were heated under pressures up to 4,000 atmospheres and the temperatures at which their magnetic properties changed were observed. No detectable change in this temperature occurred. Thus the experiments tended to confirm the view that the Earth's interior could not be ferromagnetic if it is at a high temperature. However, the experiments are not conclusive on this point since the pressures used were nearly a thousand times less than are encountered at great depths in the Earth, and great importance is still to be attached to performing the experiments at the highest pressures attainable.

In hope of obtaining some clue for the solution of problems of terrestrial magnetism an intensive program of research in nuclear physics has been pursued since 1926. While these researches have done much to enhance our knowledge of basic physics and have received wide acclaim for their contributions to the understanding of the atomic nucleus, no results so far seem *immediately* applicable to the problems of terrestrial magnetism. Yet, since pressures in the Earth's interior are so great that important atomic changes are likely to occur, theoretical developments helpful to the basic problem may be expected from more intimate knowledge of the forces within atoms.

While the contributions to our understanding of terrestrial magnetism derived from investigations in the laboratory have been negative so far their importance should not be overlooked. By the process of exhaustion such studies may lead investigators to the correct interpretation of the phenomena. The basic facts of ferromagnetism are still too poorly understood to warrant any dogmatic conclusions regarding the origin of terrestrial magnetism.

Relation of Crustal Rocks to Terrestrial Magnetism

The rocks which predominate in the Earth's crust are too feebly magnetized to account for any appreciable part of the general magnetic field. These rocks contribute some minor irregularities to the field which can be detected by careful observation. Large deposits of magnetic ores of course produce large anomalies but their effects are not perceptible at great distances. Effects produced by large masses of rocks having magnetic properties differing from those of surrounding rocks afford a basis for inferring their disposition, which has found extensive economic use in geophysical exploration.

If the magnetic field at all points on a closed surface is known a unique solution may be obtained from potential-theory for deciding what portions of the field originate within and without the bounding surface. Work done at the Department has shown that the same procedure may usually be followed if the magnetic field is known at all points on a finite plane. A unique solution is obtained under suitable conditions for deciding what portions of the field are due to causes above or below that plane. This is the method which has been used for determining what portions of the Earth's general magnetic field arise from causes beneath or above the Earth's surface. A very fundamental limitation of potential-theory prevents a unique solution for the distribution of the causes within the bounded regions—theoretically an infinite number of solutions are possible—but by consideration of *a priori* probability certain likely solutions may be deduced. By this means the configuration of magnetic rocks located at great depths can be obtained.

Most of the magnetization of crustal rocks *in situ* arises through induction produced by the Earth's general magnetic field. When these rocks are removed from their positions they exhibit a feeble residual magnetization which usually differs from that induced in them. This feeble residual magnetization may prove of great importance in the investigation of terrestrial magnetism, for it is believed that it represents the magnetization which the rocks acquired in their formation. Lavas in cooling become magnetic in the direction of the prevailing magnetic field of the Earth at the time. Similarly sediments acquire magnetization; the elementary magnetic particles distributed through them behave like tiny compasses, align themselves with the Earth's magnetic field in settling through water, and are locked in place by subsequent depositions.

Man-made records of secular variation cover a few centuries, a paltry interval for the study of a phenomenon taking place at so gradual a rate. Observations made at London extend over four centuries. There the magnetic needle pointed 11° east of north in 1580, changed gradually until it pointed 24° west of north in 1812, and since then has shifted eastward and now points only about 11° west of north. If a correct understanding of the Earth's magnetic field and its secular variation is to be attained knowledge of the limits of its variation is of utmost importance. If changes of 35° can occur in 200 years, how great may changes be which occur in thousands or millions of years? How permanent, after all, is this so-called permanent magnetic field of the Earth? In hope of answering these questions some investigators have undertaken study of the magnetization of rocks. This research has been renewed recently by the Department on a more pretentious scale than has been attempted hitherto.

Earlier investigations were concerned principally with the residual magnetization of igneous rocks, since the usually very weak magnetization of sedimentary rocks could not be accurately measured with techniques then available. Some of these igneous rocks exhibited connate magnetization in a direction opposite to that of the present magnetic field at their location, suggesting that the direction of the Earth's magnetism was completely reversed in past ages. Results obtained by measurements on igneous rocks cannot be regarded as conclusive since they may have acquired their magnetization on cooling but may have turned over a number of times before attaining their final position of repose. Accordingly, researches at the Department are directed toward the study of sedimentary rocks. To detect their weaker magnetization sensitive apparatus has been developed which enables detection of a steel magnet having a volume less than one millionth of a cubic millimeter.

Measurements have been made on the sediments taken from the ocean-bottom by the core-sampling apparatus developed by Dr. C. S. Piggot of the Institution's Geophysical Laboratory. These measurements (plate 1) have revealed changes in magnetic declination amounting to 90° during the time the sediments were deposited, presumably an interval of many thousand years. In order to secure a more detailed record similar measurements have been made on the varved Pleistocene clays left at the retreat of the last glaciation in New England. Because of the accuracy with which the varves have been dated with respect to each other they supply a comprehensive year-by-year record of fossil magnetization. Measurements made on about 100 consecutive annual varves deposited when the retreat of the glaciers began show that their magnetization is about 30° west of the present magnetic declination in that locality and that changes within this 100-year period were small, as might be expected from secular-change values in New England during the past century. No specimens so far examined suggest that the Earth's magnetism was ever completely reversed.

Continuation of these measurements throughout the entire series of dated varves, covering 7000 years of geological history, and extension of the measurements to other sediments which can be accurately dated should not only supply an extensive record of the Earth's magnetic history but may also supply a rapid means of dating deposits. Thus the study of fossil magnetism promises for the science of terrestrial magnetism what the study of organic fossils has accomplished for the science of biology.

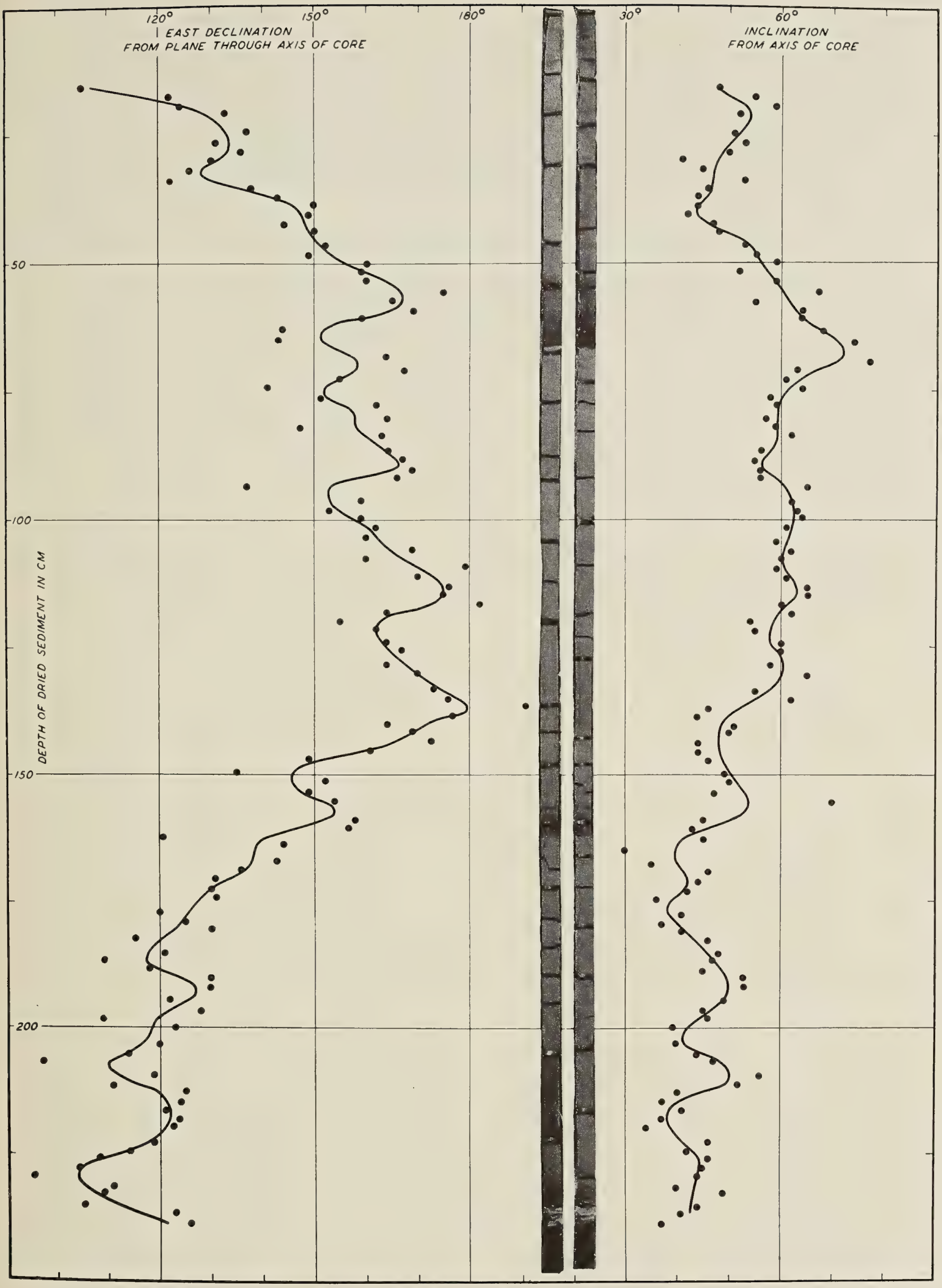
Future Research in Terrestrial Magnetism

The mode of attack on the general magnetic field of the Earth and its secular variation, as exemplified by the Department's work, consists of a continuous, planned investigation. A considerable portion of the resources available for geomagnetic research has been necessarily directed heretofore toward acquisition of data. Assumption of this phase of the investigation by governmental agencies has been fostered and is now taking place in increasing degree, permitting more emphasis on interpretative studies of data accumulated from all sources. Thus a transition from the exploratory to the interpretative stage is well under way.

Naturally we must continue looking to observations and their discussion to further understanding of the causes and laws governing geomagnetism and to provide foundations upon which tests of theory may be made. However, with the completion of a first accurate magnetic survey we may profitably utilize in geomagnetic investigation the rapid developments of high-accuracy technique in physics during the past decade, particularly nuclear physics, ionization, radioactivity, radio-wave transmission, and high-vacuum electrical research. Specific problems which were set aside years ago now may be undertaken because of the availability of these new techniques and new data.

Researches in the laboratory concerned with the basic properties of matter, especially at high pressures and magnetically and electrically, must lead to clues helpful to the understanding of the Earth's interior. Results obtainable by such specialized lines of approach are well illustrated by the contributions of the Department's ionospheric research to the solution of problems of the ephemeral variations of geomagnetism (see pages 223–247). Researches on the magnetization of crustal rocks and anomalies of the Earth's field produced by geological formations are potential sources of information but lightly drawn upon as yet.

Many of the lesser geomagnetic problems have been satisfactorily explained but the basic and greatest problem is still unsolved. Today more adequate data and new methods of investigation give encouragement to the investigator. Nevertheless the simple statement of Gilbert over three centuries ago, "Our Earth is a great magnet," offers an unwavering challenge to tax the ingenuities of modern scientific research, and calls in unique degree for international cooperative coordination of data, experiment, and discussion.



Direction of magnetization in bottom core-sample no. 3 [curve smoothed by formula $(a + 2b + 3c + 2d + e)/9$] with photograph of split core inserted to scale.

THE EPHEMERAL VARIATIONS OF THE EARTH'S MAGNETISM

L. V. BERKNER AND A. G. McNISH

Department of Terrestrial Magnetism
The Carnegie Institution of Washington

THE most remarkable characteristic of the Earth's magnetic field is its capacity for continual change. Old records show that declination of the compass at London has changed more than 36° during the past four centuries. Less lengthy series of observations at other places on the Earth reveal similar changes. Superposed upon these secular variations are the smaller ephemeral variations—more or less regular changes in direction and intensity of the Earth's field and occasional erratic fluctuations known as magnetic storms.

The cause of these ephemeral variations in the Earth's magnetism has been a subject of speculation ever since their discovery over two centuries ago. Because the daily variations are timed with respect to the position of the Sun in the sky, they were ascribed to the action of that body by even the earliest investigators. Some thought that the Sun gave rise to the variations through direct attractive action. Christie and De la Rive suggested that they were due to thermoelectric currents produced in the Earth's crust or its atmosphere through solar heating. On the basis of his experiments with the magnetic properties of oxygen, Faraday proposed that heating of the atmosphere decreased the magnetic susceptibility of its oxygen and thus influenced the magnetic field at the Earth's surface. The violent and erratic fluctuations known as magnetic storms, discovered a few years later than the diurnal variations, presented an even greater mystery.

In those pioneering days of research in terrestrial magnetism it could not be asserted to what extent the Earth's magnetism or its variations were due to causes within the Earth or in outer space. This ambiguity was removed by the genius of Gauss (see 1 under "Literature Cited" at end of paper), who showed that the Earth's field could be separated into two components by spherical harmonic analysis, one originating from causes beneath the Earth's surface and the other originating from causes in outer space. Submitting the permanent magnetic field to this mathematical treatment, he demonstrated that the predominant portion arose from internal causes. He pointed out that the procedure could be applied to the variations as well, to ascertain where they originate, though adequate data for making the application were lacking in his day.

The first spherical harmonic analysis of the diurnal variations, performed by Sir Arthur Schuster (2) toward the end of the last century, showed that the variations were due principally to causes above the Earth's surface. The portion of the diurnal variations due to internal causes was of such form and magnitude as would originate from electric currents induced in the Earth by the primary external portion. Since the appearance of Schuster's important memoir, spherical harmonic analysis has been applied to the other main ephemeral variations—the lunar diurnal variations and magnetic storms. The results have shown in all cases that the ephemeral variations are due principally to external causes, a minor portion in each case being attributable to electric currents induced within the Earth by the varying external field.

Since magnetic matter does not exist in the space immediately surrounding the Earth, the primary external portion of the ephemeral magnetic variations must be caused by electric currents, the only other known means whereby magnetic fields may be produced. Possibility of the Sun, Moon, or other celestial body which may be magnetic giving rise to the variations by direct magnetic action is precluded by the vast distances involved and by the form of the variation-field, which does not agree at all with effects which would be produced in this way. Thus the problem of accounting for the ephemeral variations in terrestrial magnetism resolves itself into determining the location, form, and origin of these electric currents flowing about the Earth. While the method introduced by Gauss permits a unique separation of the field into its internal and external portions, it supplies no definite information on the location of the causes within the internal or external regions. Such conclusions must rest upon inferences drawn from other fields of research and upon intuitive reasoning.

During recent years considerable effort has been devoted to the solution of these problems by the Department of Terrestrial Magnetism of the Carnegie Institution of Washington. Since the Earth's upper atmosphere seems a likely location for these electric currents, the electric state of those regions has been studied by radio methods. These researches have proved fruitful and have supplied valuable collateral evidence on the causes of the ephemeral variations in terrestrial magnetism. They have also revealed information of importance to other fields of science.

Although definite conclusions regarding origin may not be derived from observation and analysis of the variations at the Earth's surface, cogent inferences may be drawn from them (3). New analyses have been made and new methods of analysis have been devised so that the basic idea suggested by Gauss, nearly a century ago, is finding fresh application in the work of the Department. The purpose of this paper is to discuss these various researches and their interpretation, particular emphasis being placed on the contributions of the Department.

ORIGIN OF THE DIURNAL VARIATIONS

A comprehensive picture of what must be explained by any theory of the diurnal variations in terrestrial magnetism has been supplied by continuous photographic records of the magnetic changes made by numerous observatories widely distributed over the surface of the Earth. There are about seventy such observatories at the present time, but half of them are in the relatively small area of Europe, while extensive portions of the Earth's surface are without observatories. Thus, intimate knowledge of the magnetic variations for these extensive portions of the Earth does not exist. To ameliorate this condition, the Carnegie Institution of Washington has established and now maintains two observatories at strategic locations: one near Watheroo, western Australia, established in 1919, and another near Huancayo, Peru, established in 1922. Data from these observatories, particularly from the one near Huancayo, have been specially helpful in explaining the origin of the diurnal variations.

By application of spherical harmonic analysis to the diurnal variation as observed at many places on the Earth, a hypothetical system of electric currents flowing at some arbitrary height in the atmosphere may be mapped which would give rise to the observed magnetic variations. Several attempts have been made to account for such a system of electric currents or its equivalent.

The British physicist Balfour Stewart (4) suggested that the diurnal variations are due to electric currents induced in the atmosphere by tidal motion in the general magnetic field of the Earth. Although proposed in 1882, before it had been proved that the variations were due to external causes, this theory boldly postulated that the upper atmosphere is capable of conducting an electric current due to the action of sunlight. Researches on the electric state of the upper atmosphere by radio methods initiated in 1925 by investigators of the Department of Terrestrial Magnetism (5) and by others have since demonstrated the validity of this assumption. Thus the study of terrestrial magnetism first led to the apprehension of that region of the atmosphere, now known as the ionosphere, which is familiar to us all because it makes possible the transmission of radio waves over long distances. Because this theory of Balfour Stewart is still the most successful of all which have been proposed, and because researches at the Department of Terrestrial Magnetism have strongly supported it, requirements of the theory will be discussed in detail.

The theory proposed by Balfour Stewart depends upon the fact that the atmosphere has tides as the oceans have. These tides have been regularly recorded by barometers at a large number of stations so that the distribution of the tides over the surface of the Earth is well known. From the observed variations in barometric pressure the tidal movements of the atmosphere may be deduced. Motion of the air across the lines of force of the Earth's permanent magnetic field sets up electromotive forces in very much the same way as electromotive forces are produced in a dynamo. If the atmosphere is electrically conducting, electric currents will be caused to flow in the directions which agree very closely in form with those required to produce the diurnal magnetic variations. A similar explanation is offered for the lunar diurnal variation; in this case the lunar atmospheric tides supply the necessary motions of the air. Since the permanent magnetic field of the Earth is well known and since the air-motions are deduced from established barometric measurements, the most uncertain quantity involved in the theory is the electrical conductivity of the atmosphere.

Transmission of radio waves over long distances gave the first independent confirmation of Balfour Stewart's original assumption regarding electrification of the upper atmosphere. Subsequently verification by planned experiment still further supported the hypothesis. But this information was largely qualitative. Quantitative information on the degree of electrification of all regions of the ionosphere was required.

Early measurements on the degree of electrification of these remote regions of the atmosphere did not indicate that a sufficiently large number of electrons or ions were present to supply the electric conductivity required by the dynamo-theory as proposed by Balfour Stewart. Furthermore, careful consideration of the behavior of an ionized gas in the Earth's magnetic field showed that the large number of electrons which exist at heights above 180 km could not contribute to the electric conductivity (6). In such regions where collisions of electrons with neutral air-molecules are infrequent, the electrons will not always be driven in the directions in which they are impelled by the electromotive forces but will travel in small spirals about the lines of magnetic force, making no progress in directions perpendicular to the lines of magnetic force. Clearly, if the dynamo-theory offers the correct explanation of the diurnal variations, the electric currents must flow in the lower part of the ionosphere where frequent collisions with neutral molecules prevent the electrons from traveling in spirals.

This weakness of the dynamo-theory appeared so great that various investigators have sought other explanations for the magnetic variations. Gunn (6) proposed a diamagnetic theory based on the concept that the electrons and ions located at great heights would act as miniature magnets because of their spiraling motions and thus affect the magnetic field of the Earth's surface. Variations in the number of electrons and ions from day to night would consequently cause variations in the magnetic field. Careful analysis disclosed that magnetic variations produced in this way would be remarkably similar in their general features to the variations which are actually observed (7). Another way in which the banks of electrons and ions located at great heights might affect the diurnal variations was discussed by Chapman (8). In the combined gravitational and magnetic fields of the Earth electrons and ions might be caused to drift in such a way as to give rise to the system of electric currents required to explain the magnetic diurnal variations. As contrasted with the dynamo-theory, both of these alternate explanations depend upon the action of banks of electrons and ions at great heights. Definition of one particular region of the ionosphere as the seat of the diurnal variations would therefore supply valuable information for deciding between the theories, but such definition cannot be supplied from study of the observed variations alone.

Two independent methods of investigation have been employed to test existing theories and develop a better understanding of the genesis of magnetic variations. One of these methods seeks to ascertain the intensity of electrification of the upper atmosphere, its structure, and other physical characteristics by its ability to reflect radio waves; the other method involves detailed analysis and study of the variations themselves to compare the actual observations with those predicted by theory.

To secure information on the physical condition of the upper atmosphere, a comprehensive program of radio investigation was begun at the Department of Terrestrial Magnetism. Equipment for continuously measuring the virtual height and electrification of the various regions of the ionosphere and photographically recording those measurements has been developed at the Department (9) and installed at both of its magnetic observatories to enhance the value of magnetic observations which have been continuously recorded there. This equipment, entirely automatic in operation, transmits ten radio signals per second over a continuously varying frequency-range from 0.516 to 16.0 megacycles per second. Echoes of these signals from the reflecting regions of the ionosphere are received and recorded. The time-delays in return of the echoes indicate the virtual heights of the reflecting regions while the maximum frequency at which echoes are returned from a given region—that is, the critical frequency of penetration—is a measure of the density of electrified particles, electrons or ions, in that region. Since in many ways electrons and ions behave similarly in reflecting radio waves, an electron being about 50,000 times as effective as an ion, the terms electrons and ions are frequently used interchangeably.

Strong indications regarding the origin of the diurnal variations have been obtained from direct observation of them at the Earth's surface. Effects produced in accordance with the diamagnetic or drift-current hypothesis depend mainly on the daily variation in the number of ions in the atmosphere and on the Earth's permanent field. There is no reason for supposing that daily changes in ionic density should be very different in the Eastern and Western hemispheres. As far as they go, radio data support this view. Thus, differences in magnetic diurnal variations in the two hemispheres must be ascribed to differences in the Earth's general magnetic field. Actually, the horizontal component of the Earth's field is about 25 per cent less in equatorial regions near the 75° west meridian than it is on the other side of the Earth. Both the diamagnetic and drift-current hypotheses indicate accordingly that the diurnal variations should be about 25 per cent greater in the Western Hemisphere than in the Eastern Hemisphere.

Study of the variations at the observatory established by the Carnegie Institution near Huancayo, Peru, close to both the geographic and magnetic equators, revealed that the variations in that region are approximately three times as great as on the other side of the Earth (10). A difference as great as this does not seem reconcilable with either the diamagnetic or the drift-current hypothesis. In investigating this anomalous condition a harmonic analysis was performed at the Department (11) employing only data from stations in the Western Hemisphere. A current-system for the variations was mapped (see figure 1). This revealed an intense flow of current over the region between the geographic and magnetic equators.

Such an effect is in accordance with the dynamo-theory. Examination of the electromotive forces discloses that the currents in equatorial regions must flow against the feeble electromotive forces there, impelled by the stronger electromotive forces in higher latitudes. However, where the geographic and magnetic equators are separated, and the separation is greatest near the 75° west meridian, favorable electromotive forces are encountered around the entire circuit.

A decisive test of these theories has been supplied during the past two years by observation of radio fade-outs and magnetic effects caused by bright chromospheric eruptions on the Sun. Attention was called to these effects by J. H. Dellinger (12) of the National Bureau of Standards. Repeated observation of their occurrence has supplied the information necessary for their interpretation. Simultaneously with the visible appearance of a particularly intense bright chromospheric eruption, high-frequency radio signals cease to be reflected from the ionosphere and distinct changes take place in the Earth's magnetic field (13). These events are confined to the lighted portion of the Earth and persist for the duration of the eruption, after which normal conditions are restored (see figure 2). Because of the simultaneity of the effects on the Earth with observation of the eruption on the Sun, it is concluded that the active agent travels with the same speed as the visible light. Since ultraviolet light from the Sun under normal conditions is responsible for most of the ionization of the upper atmosphere, the inference follows that these chromospheric eruptions are special and more intense sources of the same type of radiation. That the Sun's light which gives rise to the ionized layers capable of reflecting radio waves and permitting their transmission over long distances should at times cause a cessation of these reflections is reminiscent of Æsop's fable concerning a traveler who could blow both hot and cold from the same mouth. But consideration of all the factors involved resolves this paradox into a simple picture.

In order that banks of ions may reflect radio waves, they must be located in tenuous regions so that energy absorbed from the radio waves by the ions may be re-radiated back to Earth and not wasted in random collisions with neutral molecules. Increases of ionization caused by light from the chromospheric eruptions appear to occur in levels below 90 km, the lowest level from which high-frequency waves are returned. Below the 90-km level, collisions of ions with neutral molecules are frequent so that radio waves which do not penetrate the region are absorbed, not reflected. Such low-lying absorbing regions of relatively low ion-density normally exist during the daylight hours, as is attested by the poorness of transmission on broadcast frequencies during daylight hours. Higher frequencies ordinarily penetrate this absorbing region and are reflected by the upper layers.

Striking demonstration that the solar eruptions affect mainly, and almost solely, the regions of the ionosphere below 90 km is supplied by continuous measurements of the virtual heights and ionic densities throughout the duration of an eruption (14). Actual records of a series of such measurements made by the automatic apparatus of the Department at its experimental station near Kensington, Maryland, are shown in figures *3a* and *3b*. Each frame contains a complete record through the frequency-range at 15-minute intervals from $9^{\mathrm{h}}\ 32^{\mathrm{m}}.5$ to $17^{\mathrm{h}}\ 00^{\mathrm{m}}.0$, 75° west meridian time, July 31, 1937. Time in this figure progresses from right to left. In the eighth frame the fainter echoes at the higher frequencies, to the right, have disappeared and the strong reflections at lower frequencies, to the left, have likewise faded out. No reflections are obtained on any frequency in the ninth to the thirteenth frames, after which reflections gradually reappeared, the high-frequency reflections reappearing first. Careful scaling of the records while the high-frequency reflections were recorded and low-frequency reflections were lacking reveals that the heights and ionic densities of the upper layers were not affected by the chromospheric eruption, and that the lower layer at about 100 km suffered only slight disturbance. The interpretation of these records is shown in figure 4. At the top of the figure are shown the virtual heights of the several regions of the ionosphere through the day on which the fade-out occurred compared with the averages for the week, while at the bottom the critical frequencies at which the regions are penetrated by the radio waves, and which thus indicate the maximum ionic densities of the regions, are compared with corresponding weekly averages. The lower solid curve represents the lowest wave-frequency at which reflections can be observed, a limit established by absorption in the more weakly ionized region below 90 km. During the fade-out the ability of this layer to absorb extends to higher frequencies, eventually extending to the whole range of frequencies through which reflections are ordinarily observed. The data admit of only one interpretation: the creation of dense ionization in the region below the lowest reflecting region where collisions of ions with neutral molecules are so frequent that the radio waves are absorbed.

The changes in the Earth's magnetic field associated with the fade-outs are increases in the normal diurnal variation. A representation of the ionospheric currents necessary to produce the magnetic effects associated with the radio fade-out which occurred on April 8, 1936, is represented by heavy arrows in figure 5 compared with the currents required to produce the normal diurnal variation in progress at the time of occurrence, represented by the lighter arrows. These representations derived from study of the records at a large number of observatories show that in practically every case close agreement is exhibited in the direction and relative magnitude of the two currents. The normal diurnal-variation current over Teoloyucan was anomalous on this particular day, but even this anomaly is faithfully reproduced by the special current. Thus the current-system of the special disturbance is almost a facsimile of the normal diurnal-variation current, in fact, an augmentation of it, and not an independent system which merely stimulates it.

These facts strongly indicate that the electric currents which produce the magnetic diurnal variations flow in the lower region of the ionosphere where the dynamo-theory requires that they should flow. The dynamo-theory supposes the existence of electromotive forces at all times so that enhancement of electric conductivity through increases of ionization would result in increases in current, and hence magnetic effects. Neither the diamagnetic nor the drift-current hypothesis would explain such effects as these. Thus through the powerful technique of multifrequency radio investigation of the ionosphere definite conclusions have been reached regarding the validity of theories of the magnetic diurnal variations which are in essential agreement with those obtained from study of the variations alone.

Magnetic Storms

The most spectacular ephemeral variations are magnetic storms. During the unusually severe magnetic disturbances, variations in compass-direction amounting to several degrees are noted in middle latitudes, and the intensity of the Earth's field, even in equatorial regions, may vary by as much as 4 or 5 per cent of its normal value, the changes in high latitudes being still greater. Extraordinarily brilliant flashes of auroral light are observed not only in the polar regions but well within the tropics, while serious interference is experienced in communication by radio, telephone, and telegraph over long distances.

Magnetic storms are closely associated with solar activity (15). During sunspot-maxima they are most numerous and most intense. Particularly violent storms usually occur when large, active sunspot-groups are on the surface of the Sun turned toward the Earth, although violent storms have been known to occur without visible manifestation of solar activity. Evidence of a real solar relation is well established by the fact that a pronounced 11-year cycle is revealed in magnetic activity which is more closely correlated with the sunspot-cycle than any other geophysical variable which has been extensively observed. A tendency for magnetic disturbance to recur at 27-day intervals, approximately the period of solar rotation, gives further indication of this relationship. Yet this relation of solar and terrestrial magnetic activity is only of a statistical nature; no direct correspondence between solar and terrestrial magnetic events has ever been demonstrated except the single important effect previously described—the enhancement of diurnal variation caused by bright chromospheric eruptions.

Magnetic storms, like the diurnal variations, are caused by electric phenomena taking place above the Earth's surface, a fact established by mathematical analysis. This conclusion might have been reached from consideration of the facts that auroral displays usually, if not always, accompany magnetic storms and that magnetic storms have accompanied all unusually great auroral displays. Such arguments are not always cogent. Thus, the eruption of Mount Pelée on the Island of Martinique on May 8, 1902, accompanied by a magnetic storm which began at apparently the same instant as the eruption, led to the belief that some magnetic storms originate within the Earth. Mathematical analysis of that particular storm by Bauer (16) showed that it, like all other storms which have been studied in detail, had its principal cause in electric currents flowing above the Earth's surface, so that the coincidence seems purely adventitious.

Current-systems necessary to produce magnetic storms are fairly well defined (17). For convenience of analysis the currents are divided into two systems, one consisting of a flow of electric current around the Earth with maximum density of flow in regions near the equator and also about 23° from the magnetic poles, which is uniform all around the Earth, and a second system which resembles the current-system producing the diurnal variations. During the most intense storms the total current flowing between the two poles amounts to about 10,000,000 amperes. The regions of concentrated current-flow around the poles are known as the auroral zones; they are the regions of the Earth where auroræ are most frequently seen.

The auroral zones play an important rôle in magnetic storms. Presumably they are belts in the atmosphere about 100 km above the Earth's surface, roughly elliptical in shape, about 23° from the geomagnetic poles, which are rendered highly conducting by influx of particles which give rise to auroral light. These highly conducting paths carry intense electric currents during magnetic storms. Sporadic currents which flow along these zones during magnetic disturbances amount to several million amperes in extreme cases. Since the paths are only 600 or 700 km wide, the density of flow is many times greater than the average density of flow between the two poles (see figure 6). Another indication of the importance of the auroral zones is supplied by study of the degree of magnetic disturbance in various latitudes during a storm. The maximum of disturbance occurs near the auroral zones and decreases as one moves away from the zones, either toward the poles or toward the equator.

Electrically charged corpuscles ejected from the Sun have been widely accepted as the cause of magnetic storms, but a satisfactory mechanism for their ejection from the Sun and production of their effects on the Earth has never been proposed. This corpuscular theory received its first extensive development from Birkeland and Störmer (18), later developments being due to Chapman, Ferraro (19), and others. Owing to their electric charges, these corpuscles are deflected by the Earth's permanent magnetism so that the Earth becomes engulfed in the stream, accounting for the fact that magnetic storms affect the entire Earth simultaneously. Theoretical considerations (18), supported by experiment, show that corpuscles of this sort will be precipitated with greater intensity in the regions around the poles of a magnetized sphere and give rise to auroral effects. Other corpuscles would circle about the Earth in Saturnian rings and give rise to the magnetic effects experienced in equatorial regions. Birkeland regarded the auroral-zone currents as due to these electrified particles swinging down toward the Earth's surface and out again, which, by virtue of the charges and motions of the particles, would be equivalent to electric currents.

Dissatisfaction with the corpuscular theory had led to the suggestion that magnetic storms are produced by sudden flares of ultraviolet light from the Sun (20). In accordance with this theory these flares of ultraviolet light—a pure speculation at the time they were proposed—would increase the ionization of the atmosphere and project some atoms to vast heights where they would become ionized. Traveling along the lines of the Earth's magnetic force they would fall back into the atmosphere in polar regions, thus producing auroral displays and magnetic effects. While the theory does not account for all the facts of magnetic storms, incomplete information regarding processes taking place in the high atmosphere has prevented criticism of the theory on anything more than *a priori* grounds until recently.

Weaving the various facts into a comprehensive picture of what takes place during a magnetic storm and matching this picture against theories which have been proposed to account for the phenomena is a major research of terrestrial magnetism. In order that the picture may be true and complete, new facts must be ascertained and new calculations undertaken. Much information is already available to check on many of the details of existing theories.

Observation of bright chromospheric eruptions and the terrestrial effects they produce, which has had such great significance regarding theories of the diurnal variation, has important bearing on the ultraviolet theory of magnetic storms (21). These chromospheric eruptions closely resemble the ultraviolet flares which were postulated for this theory. But ultraviolet light originating in the chromospheric eruptions has in no case given rise to a magnetic storm; it has produced effects of a distinctly different character. This argument against the ultraviolet theory is convincing though not conclusive. The character of light necessary to produce magnetic storms may be different from that originating in the eruptions, and hence the theory may still offer the correct explanation. However, this argument, though plausible, is not convincing.

Location of a definite region above the Earth's surface as the seat of the storm-currents has not yet been possible. The importance of such localization to theories of the origin of the currents has been illustrated by the conclusions drawn from this knowledge in the case of the diurnal variations.

Currents encircling the Earth in equatorial regions during magnetic storms may not be within the confines of the atmosphere, so far as past investigations have been able to show. Although no evidence which precludes their existence in the atmosphere has been revealed, still no observational demonstration has succeeded in showing that they do not flow in outer space like the rings of Saturn, a hypothesis which has been advanced. The effect of these currents is to increase the magnetic moment of the Earth, the increase lasting throughout the main phase of the storm. The total change in moment, as viewed from outer space, is dependent on the distance of the ring-current from the Earth's surface. Through work undertaken by the Institution's Committee on Coordination of Cosmic-Ray Investigations evidence having important bearing on this matter has been obtained. The intensity of cosmic radiation in low latitudes is decreased and its geographical distribution is affected by the Earth's magnetic moment. During several strong magnetic storms large decreases in cosmic-ray intensity have been reported by Forbush (22) which may be considered as the result of increases in the Earth's magnetic moment. However, these decreases are much larger than might have been expected; therefore, they are difficult to reconcile with many of our views regarding cosmic rays and magnetic storms. These effects are too novel and too involved for satisfactory explanation at present, but intensive study may elicit from them information regarding the location of these ring-currents which cannot be obtained by other means.

Attempts to determine the heights of electric currents flowing along the auroral zone have led to various results, but the most recent determinations assign them to heights not much greater than those at which the ordinary diurnal-variation currents flow (23). Birkeland (18), who first calculated the heights of these currents, obtained values ranging from about 150 km to 1000 km above the Earth's surface. Such a scattering of values cannot give a clear conception of how the currents are caused. In Birkeland's calculations no allowances were made for the effects of induced currents flowing in the Earth which might considerably alter the calculated heights. For certain practical reasons the method of spherical harmonic analysis cannot be used for separating the internal and external portions of the field-changes near the auroral zone as was done for the diurnal variations. During the past year a new method (23) of analysis has been applied in researches conducted at the Department which has accomplished this separation and led to estimates of the heights of the external currents. These calculations indicate that the currents along the auroral zone do not flow at a much greater height than the ordinary diurnal-variation currents and hence are well within the ionized regions of the atmosphere. Recent estimates of the heights based on calculations of a different sort (24) lead to the same conclusion.

Demonstration that the currents flowing along the auroral zone are confined to the Earth's atmosphere (23) is unfavorable to Birkeland's view that these currents correspond to the direct effects of charged solar corpuscles coursing through space near the Earth. Corpuscles traveling at such low heights in the ionosphere should engage in numerous collisions with the molecules in the air and thus lose their momentum. However, the view that solar corpuscles may be the indirect cause of the auroral-zone currents is still admissible. They would render the auroral region a better conductor of electric current and possibly by heating the region would give rise to winds which might produce the currents through dynamo-action. In this way their energy would be used to produce magnetic effects in a more efficient manner.

Until the recent introduction of automatic methods of registering ion-density in the ionosphere, it was practically impossible to observe the nature of fluctuations during magnetic disturbances because of the discontinuity of the earlier manually controlled methods. Although investigation of the problem by radio methods is just beginning, much information has already been obtained. The new measurements indicate (25) that the upper layers of the ionosphere, 180 km and higher, are most strongly affected during magnetic storms but this information does not demonstrate that the currents flow in those levels. The radio observations show that if the ionic density of these layers changes during a magnetic storm the change is a decrease in middle latitudes, although an increase would have been expected from *a priori* considerations. During winter day an increase is sometimes observed at the onset of very moderate storms. This apparent decrease in density does not mean that the total number of ions has decreased during a magnetic storm; the ionization may have become more diffuse. Night-time records of radio echoes show blurred reflections during magnetic disturbance suggestive of inhomogeneous layers, markedly different from the well-defined layers produced under normal conditions by ultraviolet light. Furthermore, the height increases markedly at the onset of disturbance. Such effects might be produced by corpuscles; appearance of these effects on the dark side of the Earth argues further that they are due to corpuscles in accord with the Birkeland-Störmer theory. These blurred reflections disappear during the daylight hours. Their disappearance may be due to screening of the upper layers by the better reflecting layers produced beneath by ultraviolet light or to the absorption of the blurred reflections in the low-lying layers which hinder long-distance transmission on broadcast frequencies during the day. The radio-echo data have shown that important changes in the structure of the ionosphere occur during a magnetic storm; however, a comprehensive interpretation of these changes must await the accumulation and consideration of more data.

Thus, while the last few years have seen considerable improvement in knowledge regarding magnetic storms, the investigation of these ephemeral variations has not led to such decisive results as investigations of the diurnal variations. This is due in part to the greater complexity of storm-phenomena and in part to the fact that extensive observations near the auroral zone, required for interpretation of the phenomena, cannot be readily obtained since the auroral zone traverses uninhabitable portions of the Earth.

Considerable material for the study of magnetic storms has been acquired through the International Polar Year program of 1932–1933. In this cooperative project a large number of special magnetic observatories were temporarily established in polar regions and many observatories in lower latitudes increased their program of observations. The Department of Terrestrial Magnetism took an active part in the establishment and operation of a number of these stations. Records from these stations covering a large portion of the Earth are now being assembled at places where they will be available for competent investigators. Already daily records of the magnetic changes at many of these stations for the entire thirteen months of observations have been assembled at the Department. These records constitute a more comprehensive account of the normal and idiosyncratic changes during magnetic storms than has ever before been available. It is not a vain hope that by intensive study of them supplemented by continuance of investigations of the ionosphere by radio methods a satisfactory interpretation of magnetic storms will be attained.

MAGNETIC DISTURBANCE AND RADIO TRANSMISSION

Experience has shown that most scientific researches, regardless of how abstruse or academic they may have seemed at the time, have ultimately proved of great practical importance. Little over a decade ago magnetic storms seemed to concern only scientists whose curiosity prompted them to pry into the affairs of nature. Today they have assumed a real, though frequently unapprehended, importance to most civilized beings.

A close relation exists between magnetic disturbance and radio transmission (26). During times of great magnetic disturbance transmission-conditions are so adversely affected that high-frequency transmission over long distances is practically impossible in high and middle latitudes, while in the absence of magnetic disturbance transmission is usually quite satisfactory. This applies particularly to the circuits which carry the greatest traffic, namely, those between Europe and America, because these traverse high latitudes where magnetic disturbance is most pronounced. The closeness of this relationship has been revealed by correlating radio transmission-conditions for North Atlantic circuits, measured by the transmission-disturbance figure, with magnetic activity, measured by the American magnetic character-figure for the period January 1, 1937, to April 30, 1938. The transmission-disturbance figures are daily index-numbers proportional to the average depression of signal-strength below the normal value in decibels. Of the 11 days having a magnetic character-figure greater than 1.3 (signifying a high degree of magnetic storminess) only 2 had a transmission-disturbance figure less than 4, which corresponds to a depression of the strength of the received signal in excess of 100-fold. On the other hand, of the 215 days having magnetic character-figures less than 0.3 only 5 experienced so great a depression of signal-strength. The correlation-coefficient between these two variables is 0.70, indicating that the two variables have 70 per cent of their cause in common.

Changes in the ionosphere which occur during magnetic storms suggest that this connection between magnetic activity and transmission-conditions should exist. The mechanism of this effect is not yet understood. Perhaps it arises because of destruction of the sharp boundaries of the reflecting layers so necessary for successful transmission. Possibly poor transmission in high latitudes results from formation of a low-lying absorbing layer similar in character, although different in origin, to the absorbing layer created by light from the solar flares. Either, or both, of these possibilities is suggested by the character of the records obtained during storms. It is significant that immediately upon the commencement of magnetic disturbance transmission on high frequencies is adversely affected, transmission on the highest frequencies being affected first.

Knowledge derived from observation and study of magnetic variations extending over a century may be applied directly to the problems of radio transmission because of the closeness of relationship between the two phenomena. Comprehensive data on radio transmission extend back over a single sunspot-cycle while the data on terrestrial magnetism include nearly ten complete cycles. Were it not for the extensive magnetic data the effects of solar activity on radio transmission might be regarded as purely adventitious and the relation would still be a subject for speculation, having no practical value.

Because of the close relationship between radio transmission and phenomena of terrestrial magnetism, it may be stated with assurance that conditions of transmission will be disturbed most frequently during times of sunspot-maxima and least frequently during times of sunspot-minima. The next minimum of disturbance may be expected around 1945 and the next maximum of disturbance around 1950. The present high degree of disturbance was anticipated by investigators in the field and will probably continue for a year or more with diminishing intensity.

The tendency of magnetic storms to recur at 27-day intervals is also reflected in radio conditions. In the earlier days of research in cosmic effects on radio transmission this 27-day cycle was erroneously attributed to the action of the Moon, whose revolution about the Earth requires slightly over 29 days. A similar misconception existed in terrestrial-magnetic studies ninety years ago. Knowledge gained from terrestrial magnetism has rapidly led to the association of this effect with solar rotation, active regions of the Sun which give rise to the terrestrial effects being turned toward the Earth at 27-day intervals, thus causing magnetic storms and radio disturbances.

During the past year or two when radio research has been most extensive, this 27-day recurrence-tendency has not been so marked (27) because new outbursts of solar activity have given rise to new sequences of magnetic disturbance which have disrupted the regularity of the pattern. Because of the irregularity of the 27-day pattern during the past two years, a number of investigators have expressed considerable doubt as to its reality in transmission-phenomena. However, on the basis of the magnetic data little doubt of its reality exists among magneticians and considerable assurance is felt that when the present excessive activity has subsided the 27-day pattern will be well marked as it always is during periods of less activity.

Because of this 27-day recurrence-tendency, favorable and unfavorable times for long-distance radio transmission may be forecast with a fair degree of reliability, enabling communication-companies to schedule traffic in a more advantageous manner. Considerable use has been made of this information in past years, particularly when the recurrence-tendency was most marked.

Regular weekly bulletins on the state of terrestrial-magnetic activity are now being supplied to interested individuals and organizations by the Department of Terrestrial Magnetism with the cooperation of the United States Coast and Geodetic Survey, the communication-services of the United States Army and the United States Navy, and several radio amateurs. These bulletins are based upon reports from the seven observatories operated by the Department and the Coast and Geodetic Survey. Careful study of these bulletins indicates that they are highly representative of magnetic activity for the entire Earth and bear a close relationship to radio transmission (27).

How successful forecasts of conditions of transmission will ultimately be is conjectural. Knowledge concerning the cause and mechanism of magnetic storms should improve prognostications in the same way that comprehension of the physical processes of the atmosphere is helpful to meteorological forecasts. Present knowledge of magnetic storms is largely statistical so that definite predictions similar to storm-warnings issued by the Weather Bureau cannot be made. The event or complex of events which triggers the mechanism of a magnetic storm remains unknown. But in spite of these limitations study of magnetic storms is already repaying society in practical values for the support which has been given the researches over the past century.

PHENOMENA OF THE IONOSPHERE

In attempting to solve the problems of the ephemeral variations in terrestrial magnetism, radio research has discovered a field of its own. The problems of the ionosphere have varied interests. The upper atmosphere is a great laboratory where physical and photochemical processes take place under conditions which may not be elsewhere duplicated. The absorption and dissipation of light-radiation in rare gases may be studied in an effective manner through ionospheric research. Only recently the ionosphere permitted the experimental solution of a problem in pure physics which has been argued on theoretical grounds for over a quarter of a century. The ionosphere also gives an estimate of the quality of solar radiation in the ultraviolet portion of the spectrum which cannot otherwise be observed from the Earth's surface because of absorption in the outer atmosphere. These ultraviolet radiations are much more variable than the visible radiations, which are sensibly constant, and, being intimately associated with other little-understood solar variations, give a new approach to solar physical problems. While it is not the purpose of this paper to deal thoroughly with problems of the ionosphere, a few of the more prominent advances in this field by the Department of Terrestrial Magnetism will be discussed.

There are two main regions of ionization in the upper atmosphere. When the Sun is high in the sky the upper of these regions divides into two, giving three distinct regions under these conditions. Each of these regions has its individual characteristics. The lowest of these main regions, designated as the E-region, is at a height of about 110 km and has a mean maximum density at noon corresponding to about 180,000 electrons per cubic centimeter. The next region, which exists only when the Sun is within about 45° from the zenith, is designated as the F_1-region. Its height is roughly 220 km and its mean maximum ion-density at noon is about 330,000 electrons per cubic centimeter. The highest region, designated as the F_2-region, is at a height of roughly 300 km and has a mean maximum ion-density in daylight which frequently exceeds 1,000,000 electrons per cubic centimeter [28].

The heights and densities of these layers undergo diurnal variation, indicating that radiation from the Sun is principally responsible for their existence. Variations of density in the lower layers closely follow the position of the Sun in the sky, which indicates that ultraviolet light is responsible for the effects. Variations in the highest or F_2-region are more complex, and suggest that the ionization due to ultraviolet light may be augmented by some other causes.

Variations in these layers take place throughout the year. Around the June solstice the electron-densities of the lower layers are greatest in just the amount which would be predicted from the seasonal variations in intensity of sunlight. The F_1-layer does not appear around the December solstice in latitudes as high as that of Washington.

A curious anomaly is found in the investigation of the highest or F_2-region. Simple theory of ionization by ultraviolet light from the Sun predicts that ion-density should change with altitude of the Sun, and therefore should be a simple function of declination, hour-angle, and latitude of the observing station. This is the simple type of variation normally observed for the maximum ion-densities of the lower (E- and F_1-) regions. During the winter at Washington, however, when the Sun's radiation is least intense, the F_2-region experiences its maximum average ion-density. To explain these more complex changes in the F_2-region, the concept of heating and consequent expansion of the outer atmosphere at small solar zenith-angles was introduced. This extended theory offered an explanation for the increased virtual heights of the F_2-region which are observed in summer, but failed quantitatively to predict the variation of maximum ion-density for a station in the Southern Hemisphere from the data in the Northern Hemisphere, in the following way.

Observations from the Department's observatory at Watheroo, western Australia, show that the variations in F_2-region ion-density there do not follow the pattern exhibited by the Washington data which appear typical for the Northern Hemisphere (29) with simple displacement by six months to account for the reversal of seasons.

Watheroo is located in nearly the same latitude in the Southern Hemisphere that Washington occupies in the Northern Hemisphere and, in accordance with the heating theory proposed, should experience its maximum ionization around the June solstice when it is winter there. This is not the case. A thorough analysis of the data for the F_2-region obtained by the National Bureau of Standards at Washington and by the Department of Terrestrial Magnetism at its Watheroo observatory shows that the effects may be explained as resulting from the superposition of a seasonal variation, that is, a variation which is opposite in the two hemispheres, and an annual variation which is the same in both hemispheres.

No suitable explanation for the annual variation has yet been offered. Although the Earth is closer to the Sun around the December solstice, the difference in distance is not sufficient to produce the observed effects by several orders of magnitude. The underlying significance of this new factor, therefore, is as yet unknown, and must await the collection and study of more extensive data. It emphasizes the necessity for further careful examination of the theory. The application of this annual variation-component to problems of radio transmission is most interesting, for it indicates an asymmetry of propagation of high-frequency waves in the two hemispheres, conditions in the Southern Hemisphere being the more uniform throughout the year.

A close connection between intensity of ionization and solar activity is apparent from recent observations. The general increase in ion-density observed at Washington and at Watheroo from 1934 to 1937 shows the effect of the continually increasing solar activity which occurred during the interval covered by the observations. A fivefold increase in maximum electron-density of the F_2-region has occurred from the beginning of 1934 to the end of 1937, which bespeaks an even greater increase in the ionizing power of the Sun for this region, since the ionization is thought to be proportional to the square root of the energy received. Similar but less impressive changes have occurred in the electron-density of the lower regions.

Since no appreciable change in the total visible radiation from the Sun, as received at the Earth's surface, has occurred over the same period, the great change in ionization of the F_2-region must indicate large changes in the invisible parts of the solar spectrum. Similar changes are suggested by the observations of terrestrial magnetism but the data supplied from ionospheric investigations give the more exact picture of what is occurring. Thus, from study of electrification in the upper regions of the atmosphere, information on variations in the quality of solar radiation is derived which cannot be obtained in any other way—variations which are associated with physical changes on the Sun such as sunspots and other surface-phenomena.

An interesting application of ionospheric research to the solution of a fundamental problem in pure physics has recently been made at the Department of Terrestrial Magnetism. For some years considerable uncertainty has been expressed concerning the force exerted upon an electron in a gas by an electromagnetic wave. In accordance with the theory proposed by Lorentz at the beginning of the present century, the force would be $(E + 4\pi P/3)$, where E is the electric field of the wave and the term $(4\pi P/3)$ is the force due to other electrons in the medium. It is clear that in conduction through metals the second term should not be included, and many investigators have thought that it should not be included in the case of a rare gas, so that an electron would experience only the electric force of the wave. Involved theoretical arguments have been applied to the problem and led to the conclusion that the theory as proposed by Lorentz was incorrect. No experimental demonstration had ever been achieved.

Recent experiments by the Department of Terrestrial Magnetism (30) have shown that the original theory of Lorentz as applied in the ionosphere is correct. In accordance with the theory of Lorentz, because of the second term in his expression, radio waves passing through an ionized medium in the presence of a magnetic field suffer great retardation if the frequency of the waves bears a certain relation to the magnetic field. If this term should be ignored, the retardation would occur on a quite different wave-frequency accompanied by enormously greater absorption. By transmitting waves close to the prescribed frequency the abnormal retardations have been observed in accordance with the Lorentz theory. A slight discrepancy exists between the frequency predicted from the theory of Lorentz, assuming only electrons in the medium, and the frequency at which the abnormal retardations were observed, but this disappears when the appropriate number of ions is taken into account. Inclusion of this second term in the theory of radio-wave propagation in the ionosphere has important bearing on theories of terrestrial magnetism, because it leads to values of ion-density 50 per cent greater than were previously given by the simple theory, and hence to greater values of the electric conductivity.

CONCLUSION

It would be misleading if this paper were to carry the implication that the problems of the ephemeral variations in terrestrial magnetism have been solved. Although understanding of the diurnal variations has been considerably improved during recent years, no theory regarding their origin is impregnably established. While a number of hopeful suggestions have been offered to account for the origin of magnetic storms, it is an outstanding fact that no single event or complex of events can be associated with the genesis of each of some thousand storms for which comparatively complete records exist. Because so many of these storms begin abruptly and simultaneously all over the Earth there must be some sharply defined event which produces them. What new fields of discovery and investigations may be opened up when that causal event has been isolated intrigues the imagination.

LITERATURE CITED

1. C. F. Gauss, Allgemeine Theorie des Erdmagnetismus, in *Resultate aus den Beobachtungen des Magnetischen Vereins im Jahre 1838.* Leipzig, 1839.
2. A. Schuster, *Phil. Trans. R. Soc.*, vol. 180, pp. 467–519, 1889.
3. S. Chapman, *Phil. Trans. R. Soc.*, vol. 218, pp. 1–118, 1919.
4. Balfour Stewart, *Encycl. Brit.*, 9th ed., vol. 16, 159–189, 1882.
5. G. Breit and M. A. Tuve, *Phys. Rev.*, vol. 28, pp. 554–575, 1926; A. H. Taylor and E. O. Hulburt, *Phys. Rev.*, vol. 27, pp. 189–215, 1926; E. V. Appleton, *Proc. Phys. Soc., London*, vol. 37, pp. 160–220, 1925.
6. R. Gunn, *Phys. Rev.*, vol. 32, pp. 133–141, 1928.
7. S. Chapman, *Terr. Mag.*, vol. 34, pp. 1–16, 1929.
8. S. Chapman, *Proc. R. Soc.*, vol. 122, pp. 369–386, 1929.
9. L. V. Berkner, H. W. Wells, and S. L. Seaton, *Union Geod. Geophys. Inter nat., Sect. Mag. Elect. Terr., Bull. No. 10*, pp. 340–357, 1937.
10. H. F. Johnston and A. G. McNish, *C. R. Cong. Internat. Electricité, Paris*, vol. 12, pp. 41–52, 1932.
11. A. G. McNish, *Union Geod. Geophys. Internat., Sect. Mag. Elect. Terr., Bull. No. 10*, 271–280, 1937.
12. J. H. Dellinger, *Phys. Rev.*, vol. 48, p. 705, 1935; vol. 50, p. 1189, 1936.

13. J. A. Fleming, *Terr. Mag.*, vol. 41, pp. 404–406, 1936; A. G. McNish, *Terr. Mag.*, vol. 42, pp. 109–122, 1937.
14. L. V. Berkner and H. W. Wells, *Terr. Mag.*, vol. 42, pp. 301–309, 1937.
15. J. Bartels, *Terr. Mag.*, vol. 37, pp. 1–52, 1932.
16. L. A. Bauer, *Terr. Mag.*, vol. 15, pp. 9–20, 1910, and unpublished work.
17. S. Chapman, *Terr. Mag.*, vol. 40, pp. 349–370, 1935.
18. Kr. Birkeland and C. Störmer, The Norwegian Aurora Polaris Expedition, 1902–1903. Christiania, 1913.
19. S. Chapman and V. C. A. Ferraro, *Terr. Mag.*, vol. 36, pp. 77–97, 171–186, 1931; vol. 37, pp. 147–156, 1932; V. C. A. Ferraro, *Terr. Mag.*, vol. 37, pp. 421–429, 1932; S. Chapman and V. C. A. Ferraro, *Terr. Mag.*, vol. 38, pp. 79–96, 1933.
20. H. B. Maris and E. O. Hulburt, *Phys. Rev.*, vol. 33, pp. 412–431, 1929; E. O. Hulburt, *Phys. Rev.*, vol. 34, pp. 344–351, 1929; vol. 36, pp. 1560–1569, 1930.
21. A. G. McNish, *Phys. Rev.*, vol. 52, pp. 155–160, 1937.
22. S. E. Forbush, *Phys. Rev.*, vol. 51, pp. 1108–1109, 1937.
23. A. G. McNish, *Terr. Mag.*, vol. 43, pp. 67–75, 1938.
24. S. Chapman and E. H. Vestine, unpublished investigations.
25. L. V. Berkner, *Trans. Amer. Geophys. Union, 19th Ann. Meeting*, 1938.
26. A. G. McNish and A. K. Ludy, *Terr. Mag.*, vol. 42, pp. 173–177, 1937.
27. A. G. McNish and H. F. Johnston, *Terr. Mag.*, vol. 43, December, 1938.
28. L. V. Berkner, *Sci. Mon.*, vol. 45, pp. 126–141, 1936.
29. L. V. Berkner and H. W. Wells, *Terr. Mag.*, vol. 43, pp. 15–40, 1938.
30. H. G. Booker and L. V. Berkner, *Nature*, vol. 141, pp. 562–563, 1938.

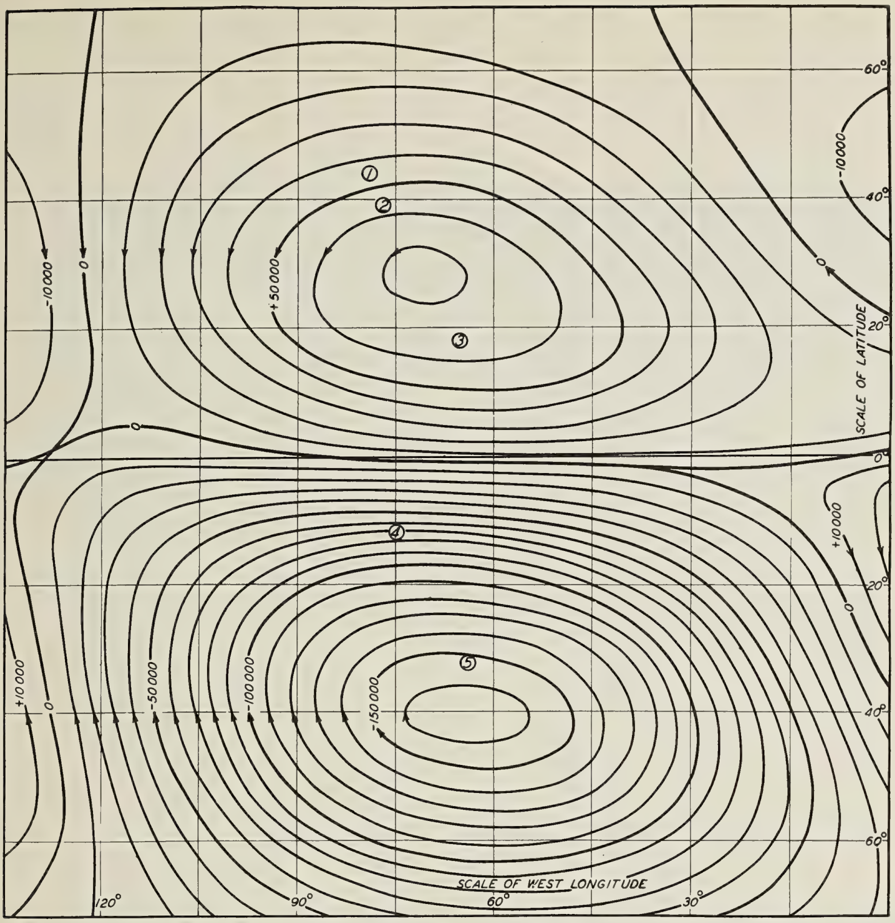
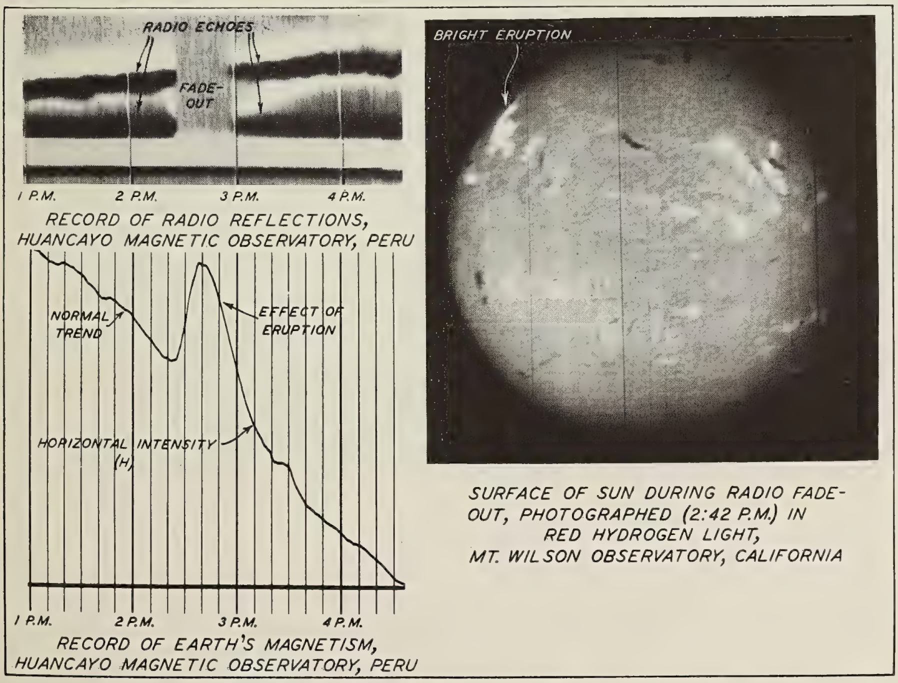

60°
40°
20°
0°
20°
40°
60°
-10000
0
+50000
+10000
-50000
-100000
-150000
①
②
③
④
⑤
SCALE OF LATITUDE
SCALE OF WEST LONGITUDE
120°
90°
60°
30°

Fig. 1. Electric currents in amperes flowing in ionosphere required to cause diurnal variations during daylight in Western Hemisphere at Agincourt (1), Cheltenham (2), Vieques (3), Huancayo (4), and Pilar (5).



RECORD OF RADIO REFLECTIONS, HUANCAYO MAGNETIC OBSERVATORY, PERU

RECORD OF EARTH'S MAGNETISM, HUANCAYO MAGNETIC OBSERVATORY, PERU

SURFACE OF SUN DURING RADIO FADE-OUT, PHOTOGRAPHED (2:42 P.M.) IN RED HYDROGEN LIGHT, MT. WILSON OBSERVATORY, CALIFORNIA

Fig. 2. Changes in Earth's magnetism and cessation of radio reflections from ionosphere accompanying bright eruption in solar chromosphere, August 28, 1937.

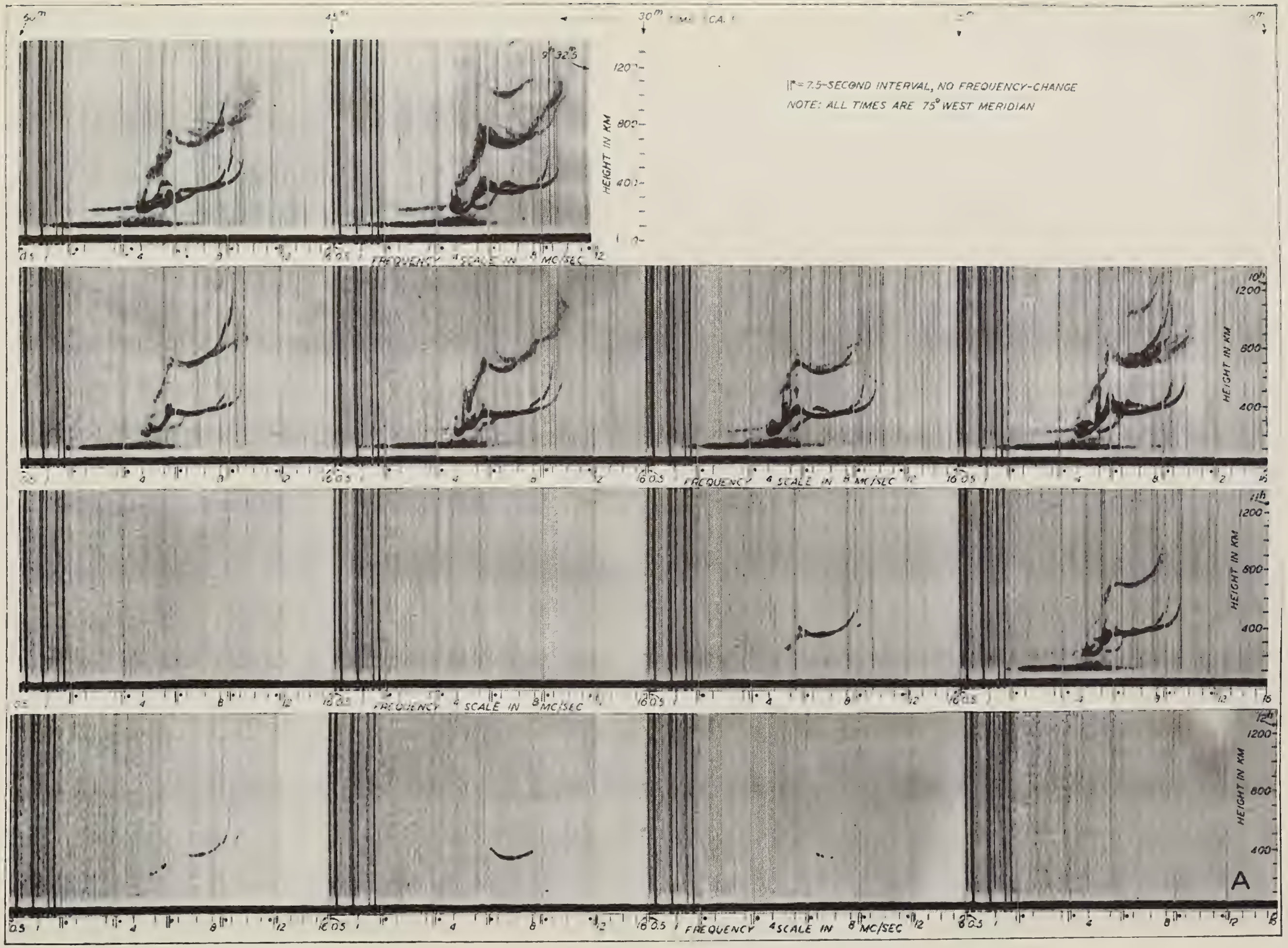


Fig. 3*a*. Ionospheric virtual heights from multifrequency records, Kensington, Maryland, July 31, 1937, for 0.516 to 16.0 mc/sec over 15-minute periods ($9^h\ 32^m.5$ to $13^h\ 00^m$, 75° west meridian mean time), showing effect of fade-out.

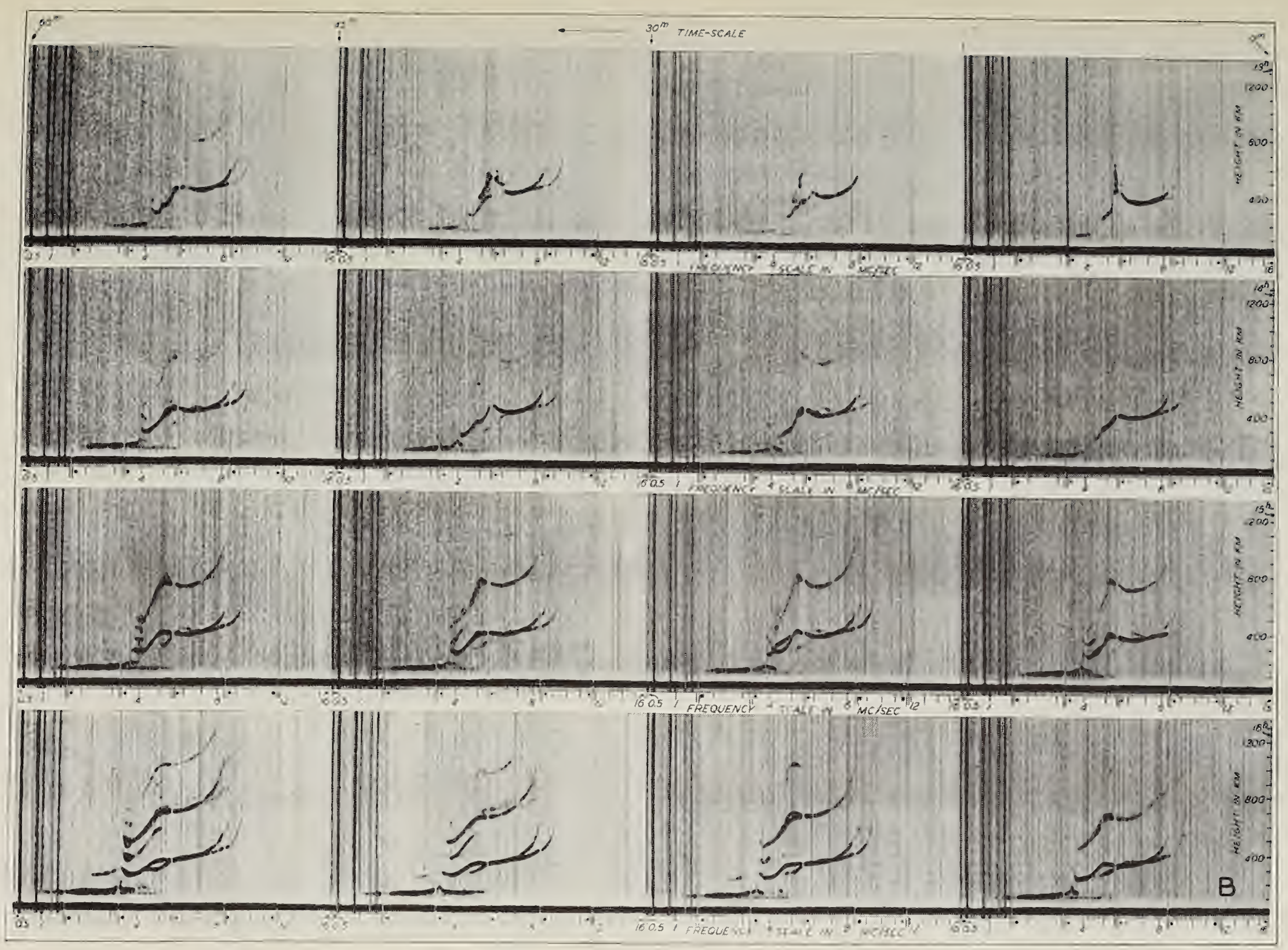


Fig. 3*b*. Ionospheric virtual heights from multifrequency records, Kensington, Maryland, July 31, 1937, for 0.516 to 16.0 mc/sec over 15-minute periods (13^h 00^m to 17^h 00^m, 75° west meridian mean time), showing effect of fade-out.

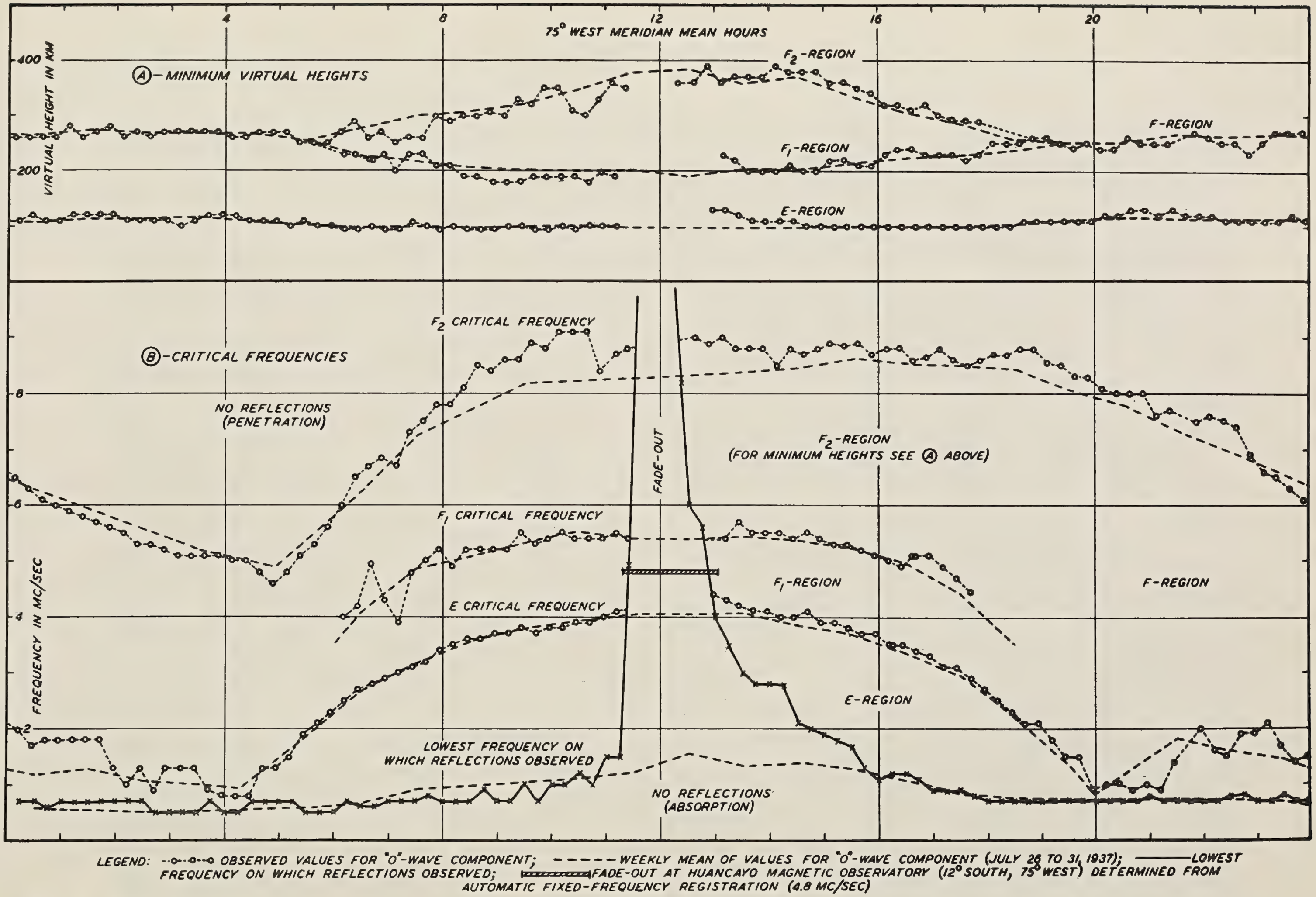


Fig. 4. The ionosphere showing radio fade-out July 31, 1937, determined from automatic multifrequency registrations, Kensington, Maryland.

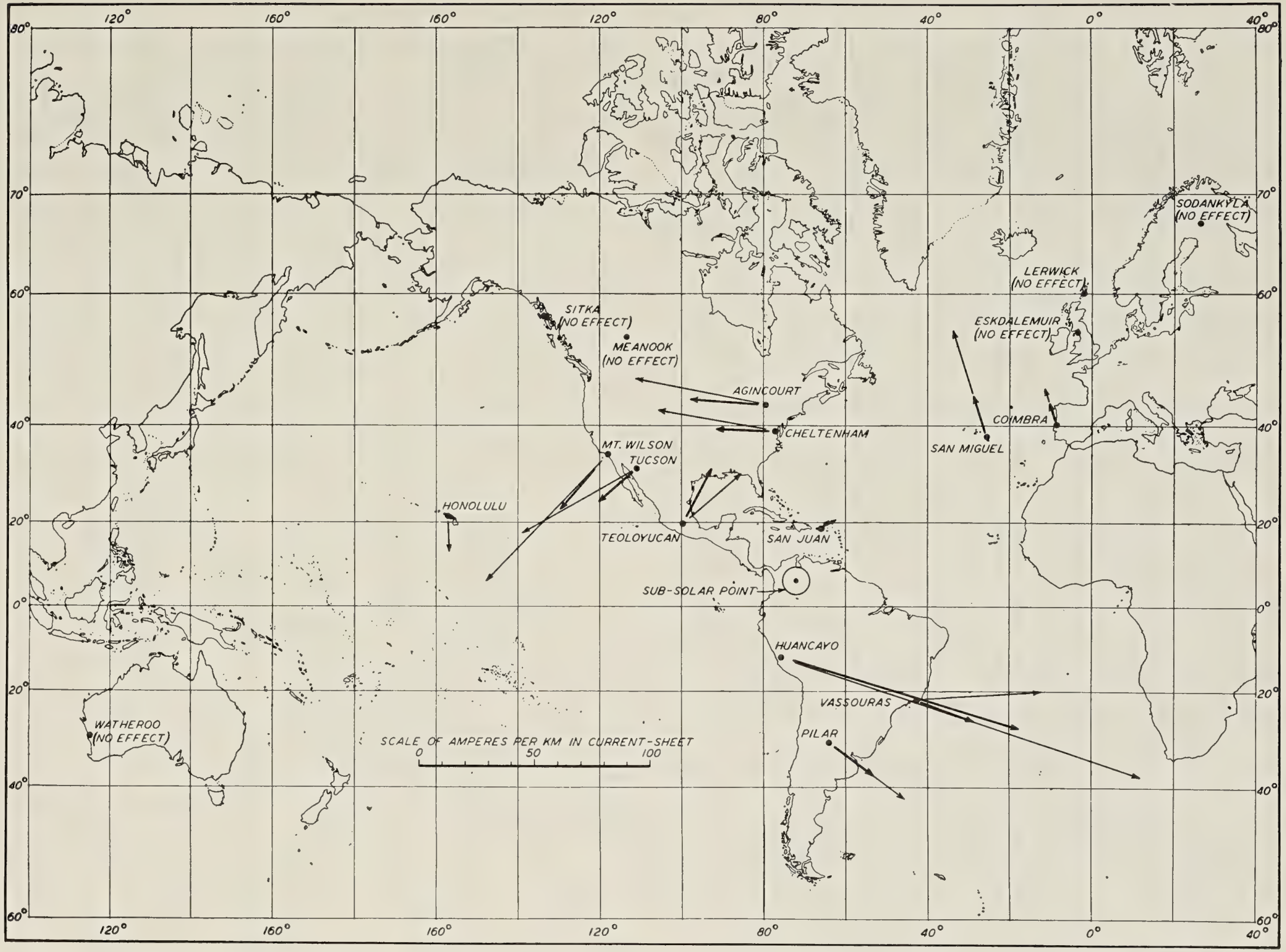

120°
160°
160°
120°
80°
40°
0°
40°
80°
70°
60°
40°
20°
0°
20°
40°
60°
SODANKYLÄ
(NO EFFECT)
LERWICK
(NO EFFECT)
ESKDALEMUIR
(NO EFFECT)
COIMBRA
SAN MIGUEL
SITKA
(NO EFFECT)
MEANOOK
(NO EFFECT)
AGINCOURT
CHELTENHAM
MT. WILSON
TUCSON
HONOLULU
TEOLOYUCAN
SAN JUAN
SUB-SOLAR POINT
HUANCAYO
VASSOURAS
PILAR
WATHEROO
(NO EFFECT)
SCALE OF AMPERES PER KM IN CURRENT-SHEET
0
50
100

Fig. 5. Overhead currents necessary to produce magnetic changes at time of solar flare beginning at 16^h 46^m GMT, April 8, 1936.

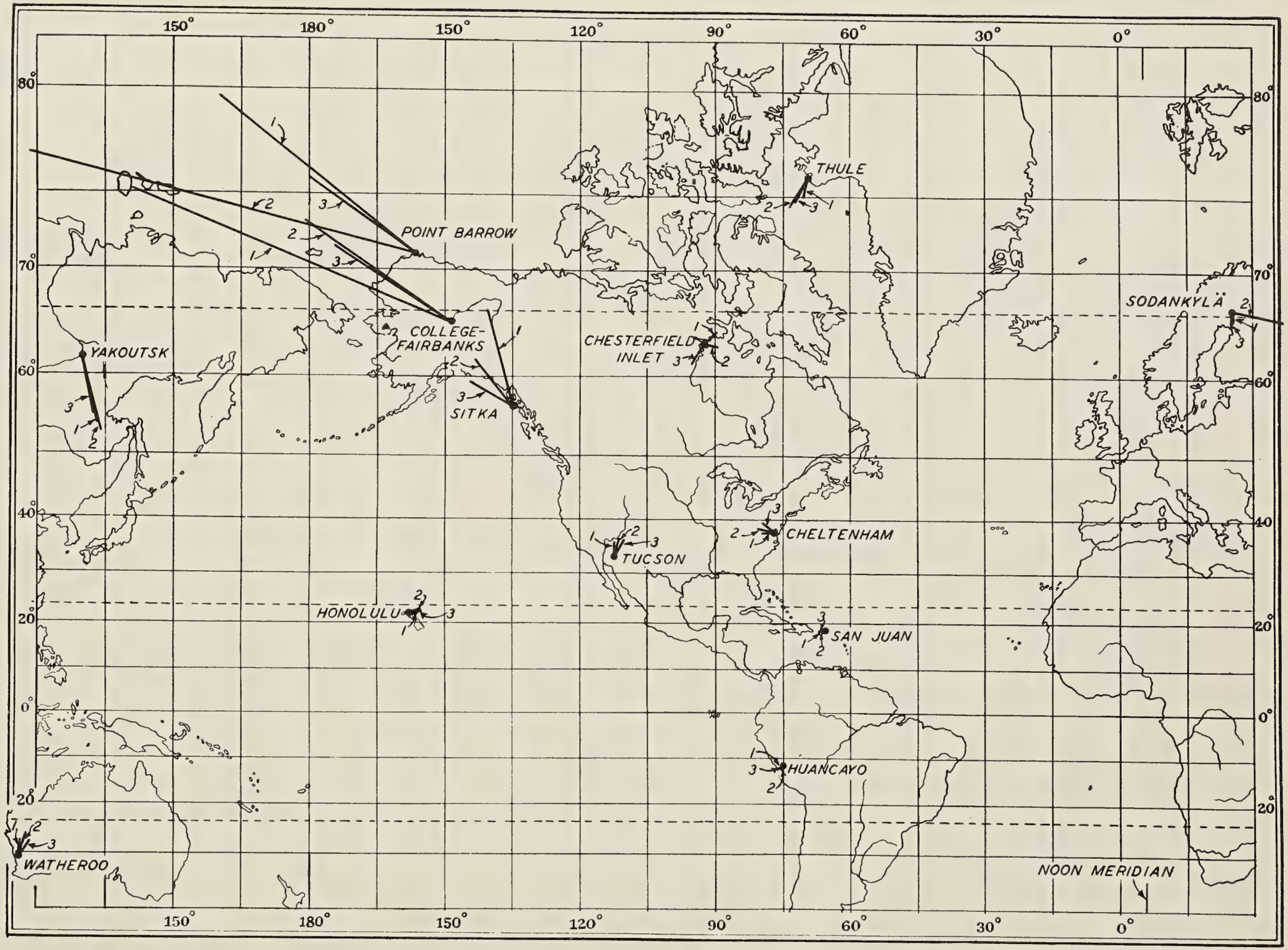

150°
180°
150°
120°
90°
60°
30°
0°
80°
70°
60°
40°
20°
0°
20°
POINT BARROW
THULE
SODANKYLÄ
COLLEGE-FAIRBANKS
CHESTERFIELD INLET
YAKOUTSK
SITKA
CHELTENHAM
TUCSON
HONOLULU
SAN JUAN
HUANCAYO
WATHEROO
NOON MERIDIAN
1
2
3

Fig. 6. Directions and relative magnitudes of currents necessary to produce magnetic bay, April 3, 1933, referred to terminal values at $11^h\ 10^m$ and $12^h\ 30^m$ GMT: (1) at $11^h\ 25^m$; (2) at $11^h\ 40^m$; (3) at $11^h\ 55^m$.

BIOLOGICAL SCIENCES

THE COOPERATIVE RESEARCHES OF THE NUTRITION LABORATORY

FRANCIS G. BENEDICT

Research Associate, The Carnegie Institution of Washington
Formerly Director, Nutrition Laboratory

With the present-day specialization in scientific research, stress is being laid by workers on the ramifications of research paths radiating with widening divergence from one major topic. The accumulation of information in any field is primarily the function of the systematist; and while the ever sharpening focus on the picture of the field remains, the wider association of the picture with other problems in the same branch of science and with the general development of knowledge is often obscured. In a homely way, one may state that he does not see the forest for the trees. A proper perspective is often denied the intensive worker, and it is only when he comes in contact with students or workers in other fields that he grasps the broader concept. The cloistered scientist has performed his work; by intensity of application, but often with exclusive concentration, he has brought science from a plaything or a curiosity to the state of exact, reliable information—information that has passed into the general knowledge of man and thereafter, directly or indirectly, has affected his livelihood and habit of thought. Today the systematist, with his extreme care, his oblivious seclusion, has, to be sure, not lost his usefulness, for there will always be need for the widening of basic knowledge; but no longer may a scientist consider that his obligation is solely to his own field. Scientists who have cultivated external interests realize that the benefit derived therefrom is always manifold. The wider usefulness or importance of their own findings to life in general is more patent. Their observations stimulate to further endeavor, and frequently bring side lights from an entirely different angle that illuminate the problem under investigation in its narrowest sense.

The individual research worker, with modest means, laboratory resources, and assistance, might think that great research organizations would find complete and full contacts in the individual personnel of each research laboratory. This is, to be sure, a great advance on the cloistered worker, but major problems almost invariably overtax the most complete research organizations and cooperative endeavor is imperatively needed. The Nutrition Laboratory, with its liberal equipment, its subsidy, and highly trained staff, only too often sensed an incompleteness in the forces assembled in attacking problems, and early made use of cooperative and willing colleagues.

In a new field, such as the study of the laws governing vital activity in humans—our basic thesis—the development of experimental procedures necessary for a complete survey of men and animals occupied such a large portion of our time that the application of these techniques, the important accumulation of experimental material, taxed our resources too heavily. Neither humans nor animals were available in sufficient numbers, in the confines of our laboratory, to insure a sufficiently rapid and extensive survey of these groups with the necessarily complicated and time-consuming technique. External contacts were indispensable. As the material began to be accumulated, with its emphasis on humans, the resolving of the data presented special problems. Nowhere had these expensive, complicated metabolism surveys of normal humans been approximated, and, exactly as the collection of data involved highly special techniques, we early saw that possibly simple averaging and charting might not bring out with clarity many conclusions that such a general survey should supply.

Thus early we interested a specialist within the Institution in the person of the late Dr. J. Arthur Harris, whose skill and ingenuity in mathematical treatment of large amounts of biological material made biometry a new tool of incalculable service, a tool that we would have wholly missed had we relied on our own myopic vision. Again, with the rapid introduction of biometric methods (many think altogether too rapid an advancement), the Nutrition Laboratory, in consultation with Professor E. B. Wilson, sought to avoid the all-too-current, unjustifiable mathematical orgies indulged in by many writers in applying biometric methods.

Biological measures may be made readily on hundreds of small animals, mice, rats, guinea pigs, or rabbits, but with the larger animals such a mass of experimental material is not available. For a survey of the metabolism (heat production) of monkeys a monkey colony is imperative, of chimpanzees a zoological garden with a large colony or, better still, a special laboratory where large numbers of chimpanzees are available. Studies with sheep, cows, and horses obviously need contacts with great agricultural organizations. Yet a reasonably complete survey of precisely these species was necessary to our main thesis. No study of vital energetics could be complete without a knowledge of the heat production of the cold-blooded animals, and while frogs and small fish might be studied in the laboratory confines, a more extensive survey is essential; just as the large warm-blooded animals called for special conditions, colonies, etc., impossible to maintain at the Nutrition Laboratory, we must have access to a fine collection of cold-blooded animals, such as amphibians, snakes, and tortoises. It is hardly the place to list the innumerable laboratories and research workers drawn upon for this program. Cooperation of the most intense order alone made the program possible.

In human pathology it was early seen that disturbances of the thyroid, especially in goiter and myxedema, produced profound changes in the metabolism from normal, an overactive thyroid, as in toxic goiter, giving at times a doubling of the metabolism, while a deficiency of thyroid produced a profound lowering. Obviously any research in these two instances of perturbed metabolism required connection with a hospital, and the resources of the Joslin Clinic and the New England Deaconess Hospital were rapidly drawn upon. Our early study of diabetes led to the extensive use of Nutrition Laboratory techniques in interpreting the complicated picture presented by diabetes, with its then prevalent accompanying acidosis. Quite aside from the light such studies threw on diabetes, they have proved of incalculable value in illuminating the various factors affecting the level of vital activity in normal humans.

The profound alterations accompanying different thyroid activities soon made the general use of basal metabolism measures imperatively necessary in all hospitals in diagnosing these cases and in following treatment. The sweeping use of this technique, soon accepted by all hospitals, closed the Nutrition Laboratory's active participation in such studies. With the hospital acceptance, introduction, and general use of these techniques, studies in pathological metabolism naturally followed, and today no hospital is considered well equipped without one or more of the numerous forms or modifications of basal metabolism apparatus.

A most striking illustration of the effectiveness of cooperative research in which large research organizations or institutions are involved, rather than perhaps simply individuals, was brought out in connection with the Nutrition Laboratory's survey of human metabolism, with particular regard to race. For decades a physiological concept had obtained that the basic heat production of all warm-blooded animals was controlled by the surface area and that the heat production was, roughly speaking, 1000 calories per 24 hours per square meter of surface area. This concept, originally founded on experiments made on dogs of various sizes, was considered as obtaining likewise with other animals, specifically humans. This belief has become so fixed in the minds of a very large number of research workers that today it is held by most as one of the basic concepts in human physiology. A simpler unifying principle, that measures vital activity not only of humans but of other animals, could hardly be conceived. In its very extensive survey of the basal metabolism of humans, the Nutrition Laboratory early found very considerable deviations from this concept and frequently challenged the true value of the so-called "law of surface area."

Do men differ in their basic vital activity, their basal metabolism? To what extent do they differ, and above all, are differences in metabolism associated with pronounced dietetic characteristics or composition of the body, such as the proportion of body fat, etc.? With reasonable certainty that the heat production of humans was not controlled by size, the differences assumed an unusual importance. Above all, why do these differences exist?

In the cosmopolitan population of a large city, great variations in physical configuration are seen, but groups of humans are more differentiated by racial characteristics in general, such as weight, height, configuration, and color, nature of the diet, and above all, temperament.

To find persons available for study in sufficient numbers and conditioned for metabolism experiments is out of the question, even in large cosmopolitan centers. Races of persons must be studied in their own habitat, where climate, food, and physical activity are more uniform; and one seeks to learn: (*a*) Have all men of a given race uniform metabolism? (*b*) Do races differ in their heat production? This is the most perfect measure or yardstick of those complicated pooled activities listed under the head of vital activity.

In a cooperative research with Mount Holyoke College, it so happened that a number of Oriental women students had their basal metabolism measured in connection with an entirely different problem. Here the strikingly low metabolism noted led to the conviction that a general survey of the races must be made. Although many of our large metropolitan centers have local racial groups, it is obvious that the problem could only be satisfactorily studied by a field survey in the several countries.

A hint as to the low metabolism of the Chinese had been given by Dr. H. G. Earle a number of years before. There then began a series of cooperative racial studies which rapidly grew in extent and number, finally including several sections of China, India, Australia, Hawaii, Samoa, the Maya in Yucatan, and the browns and blacks in Jamaica. Obviously, surveys as extensive as these must involve cooperation, and our efforts at securing such varied contacts, with at times actual field studies involving almost insuperable difficulties, is a most interesting chapter in the story of the success and value of cooperative enterprises. Obviously, without such cooperation the Nutrition Laboratory, even by sending members of its own staff into the field, could have secured but a fraction of the data needed for an adequate study of this important point, the influence of race upon human basal metabolism.

A particular value of the Nutrition Laboratory's survey, however, lies in the fact that when the final comparison of all the data was made, one had to deal with a series of observations made by a single accurately tested technique and by observers who had either received their training in the Nutrition Laboratory or been trained by Nutrition Laboratory representatives. Finally, the greater number of the research experiments were made with a single type of apparatus loaned by the Nutrition Laboratory to the cooperating investigator. It was believed that in this wise the variable difficulties in technique could be for the most part eliminated, and that any differences in basal metabolism found as a result of the study could properly be ascribed to causes other than errors in technique or individuality of the operator. As an indication of the extent of these cooperative investigations, a list of our associates follows:

Dr. Eleanor Mason studied intensively in Madras the basal metabolism of groups of native Indian women, particularly the Tamils and the Malayalis, in both of which groups a strikingly low metabolism was found.

Dean Stanley D. Wilson, together with several associates, devoted the greater part of two years of his time to a study of the Chinese students in Peiping.

Dr. H. S. D. Garven studied a number of groups at Mukden, and Professor Kilborn made several difficult field excursions in Szechwan.

In Australia Sir C. Stanton Hicks studied the bush natives, again under almost insuperable difficulties.

Professor Carey D. Miller made an exhaustive study of the several races and racial mixtures in Hawaii, culminating in a study of Polynesians at Pango Pango, American Samoa.

In the Western Hemisphere, Dr. Morris Steggerda surveyed the browns and blacks of Jamaica.

In connection with other Carnegie Institution expeditions, it was possible for us to make an especially intensive survey of the Maya in Yucatan. In the first expedition, Professor George D. Williams noted a significantly high metabolism among these people. This was so challenging that on a second expedition Dr. George C. Shattuck, repeating this work, confirmed it fully. Realizing the physiological importance of this find, a still further study was carried on through a third expedition in which Dr. Steggerda, with his long experience in these studies, again surveyed a group. It is gratifying to note that all three surveys were in complete accord, indicating a significantly high metabolism in this group.

Although the Laboratory itself did not join directly in the cooperative research of Dr. Pi-Suñer among the natives in Chili, nevertheless Dr. Pi-Suñer had obtained his training in metabolism at the Nutrition Laboratory and very kindly placed his results at our disposal. Araucanian Mapuches were studied and gave a noticeably high basal metabolic rate. The finding of high basal metabolism in these natives and in the Maya was further supplemented by findings with regard to a Miao race studied by Dr. Garven in

West China. The significance of the entire survey, however, is in the fact that, although it has been assumed that the heat production of humans is extremely uniform per unit of surface area, this survey has shown that within the human species the metabolism of a group, not an individual, may vary from the normal standard by from − 17 per cent to + 10 or even + 15 per cent.

The Nutrition Laboratory, in its finely equipped building and with its unusually well-trained staff, could not possibly have consummated these researches without the active participation of these numerous institutes and cooperating investigators. On the other hand, the Nutrition Laboratory has, by such cooperation, been instrumental in giving to the cooperating investigators information, obtained with authoritative, tested, and uniform technique, regarding the particular metabolic activity of each special race. The survey in many instances was able to rule out differences in diet, climate, activity, etc., and finally ascribe the major part of differences in metabolism found to purely *racial* characteristics. For example, certain Chinese were studied in Boston, and they had precisely the same characteristic low metabolism as the Chinese in China. These racial studies on humans, showing clearly established differences in metabolism, call for further studies and for a discussion and explanation of the metabolic differences. This latter factor cannot be handled by the research worker sitting in the laboratory; he must have cooperation with the field worker and the person who collected the data.

The Nutrition Laboratory has infused its techniques, its concepts, and its findings into innumerable clinics and hospitals in dealing with humans, as well as into animal nutrition, especially of large domestic animals, and thus has attempted to repay in slight measure the debt it owes to its cooperating investigators. These studies have been at times long and arduous, at times drab, but basically valuable. The foundation for knowledge of the normal metabolism of many species of animals and many races of men has been solidly laid, all of which would have been wholly impossible in the confines of a single laboratory, no matter how munificently subsidized, and is a real product of the fullest cooperation with other institutes of research.

Emphasis may be laid on the fact that, while the individual in the university department readily calls upon his associates in his department for aid and cooperation with resources that are often extremely limited, yet even large research units with seemingly full equipment and diversified staff are of necessity engaged in major problems that tax to the utmost their abundant resources and hence imperatively seek cooperation for fulfilment of plans. So cooperation is as essential to the large, well-equipped institution as to these shoulder-to-shoulder efforts of members of a small university department. The contacts and affiliations are different only in degree. The relatively simple program of the isolated worker needs for its fulfilment friendly adverse criticism and not infrequently joint effort. The large research groups are occupied with major problems and programs that cannot be completed without intimate contact with other cooperative groups. Only in this way can the fullest development of a research program be brought into being. A further, often entirely overlooked, outcome of such cooperative effort is that in the resolution of data a wider vision obtains and intimate prepublication discussion, with its clarifying influences, takes the place of trouble-making, usually profitless, polemical tirades.

THE CHANGING ORGANISM

OSCAR RIDDLE

Department of Genetics
The Carnegie Institution of Washington

Introduction

In most biological investigation the center of the stage is firmly held by the *living* organism, and good lights should focus constantly on that active entity. Under such lights and perspectives as are available to us it nevertheless seems that to different eyes the living organism is different things; in other words, the organism is variously regarded as stable and as unstable. From any point of view it is clear that during its development from egg to adult a higher organism undergoes amazing and almost limitless change, albeit certain physical and chemical characteristics of the blood (the internal environment) even then show relative constancy. Nevertheless by different biological disciplines the adult or "finished" warm-blooded animal is commonly credited with quite unlike degrees of stability.

One or two illustrations of this difference may be cited. The evolutionist and geneticist, for example, are impressed with the persistence and relative immutability of whatever type may occupy the center of their stage; geneticists are especially concerned with the share of the gene in building an organism of particular and predictable type and function, and with possibilities of combining or of changing genes for the production of other new but relatively stable types. On the other hand, the psychologist is much impressed with endless additions of new background and machinery of mind during all postnatal life; here it is clear that somatic change is reflected in the mental sphere, and progress in certain limited aspects of psychological study necessarily awaits or attends a disclosure and appraisal of the biological changes which actually occur within the organism. Again, physiologists and biochemists, though quite familiar with very many such things as the irregular tides of bodily and cellular activity, of nutritive state, and of adjustment to changes in external and internal environment, have usually seemed to be most impressed hitherto by the presence in man and other higher organisms of mechanisms for preserving a "steady state" (homeostasis) in such basic matters as hydrogen-ion of blood, body temperature, and basal metabolic rate.

That these and still other impressive elements of stability actually exist in all organisms, higher and lower, is fully granted here. In this paper, however, the writer wishes to outline some investigations (largely conducted by himself and associates, and in part in collaboration with other departments of Carnegie Institution) which give emphasis to the view that even in the adults of higher animals our newer knowledge discloses elements of flowing bodily change—involving marked change in parts of the internal environment—which characterize the organism perhaps more truly than do the several admitted elements of stability. The preparation of this outline is prompted by the view that some part of the success of many present and future biological studies would seem to depend upon very full recognition of the fact that we make our measurements not merely upon an organism but upon a *changing* organism.

Only a brief survey of this subject is attempted here. Indeed, discussion and illustration are limited to certain changes readily considered under the following headings: growth, seasons, the reproductive cycle, ovulation, daily rhythms. Again, our survey of these categories of change will be restricted essentially to changes observed in the pigeon. Nearly all of the items included here have been published earlier and in a much more complete form. Though we have briefly stressed the concept of "the changing organism" in some of those studies (1, 2; numbers refer to list of literature cited, at end of paper) there has been no earlier opportunity to draw attention to the several and profound aspects of change which attend the life cycle of an individual of this species.

Changes during Growth

Something of the rapidity of change during growth, and something of the individuality and unpredictability of certain aspects of growth, as these have been measured in a single race of pigeons, may be seen in figure 1. The various curves employ broken lines to bridge the gap between values obtained on the young and other values obtained on truly adult (439-day-old) birds. The growth of the body as a whole is extremely rapid in all races of pigeons (3, 4).

Two remarkable organs share this phase of rapid growth (to 80–90 days) in the body as a whole, only to begin a rapid decrease of weight at or near the 80-day stage: the thymus loses approximately 50 per cent of its weight within the next 3 months (5), and within a slightly longer period the bursa Fabricii usually practically disappears (3). Growth changes in the testis or ovary of the pigeon contrast most strongly with the two organs just mentioned. It is only when thymus and bursa have reached maximum size or have begun to involute, and when adult body size has been attained, that the germ glands begin notable growth (3). The period of repressed growth in the testis, and also its spurt of growth after 80 days of life, have been at least partially associated with a lack of production and release of follicle-stimulating hormone in the pigeon pituitary during the first two or three months of life and with a subsequent adequate production of this hormone (6, 7).

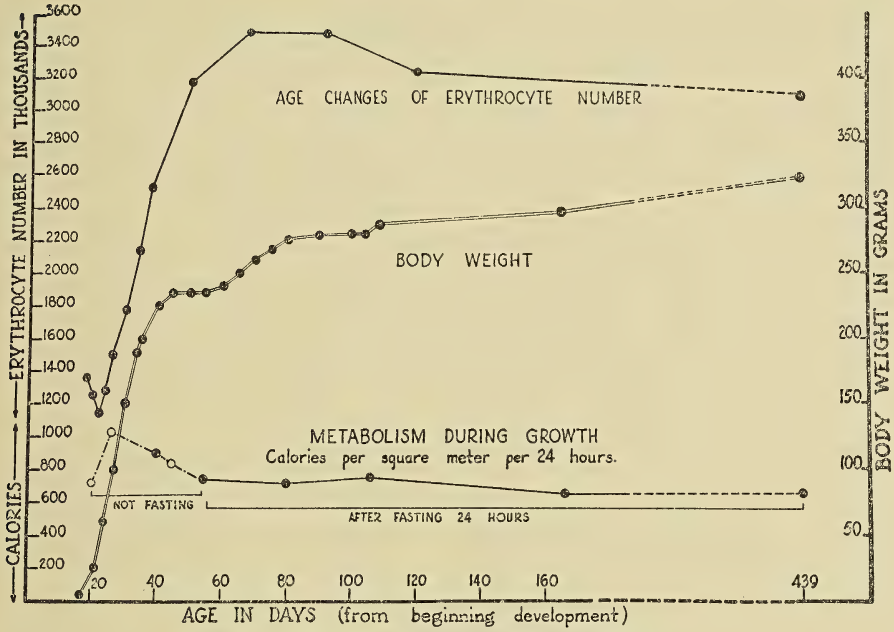


FIG. 1. Changes during growth in Tippler pigeons.

The changes in erythrocyte number during growth, as figured here for one race of pigeons, was observed in all of four races studied (8). It is particularly notable that during the first three days of postnatal life the number of red cells decreases from 1,275,200 to 1,078,000 per cubic millimeter. About 27 days later these red cells attain a maximum value (3,409,700), which in turn is maintained for only about 30 to 35 days. This remarkable reduction of red cell number during the first days of postnatal life is as yet unexplained, but it has been observed in rat, rabbit, man, and pigeon.

The basal metabolic rate in Tippler pigeons seems by no means to be an unchanging value (9) during the period of rapid body growth. It is true that the four groups of measurements (open circles on graph) which show precisely these aberrant and changing values were for the most part not made on fasting animals (the values shown are based upon a correction factor) and that some uncertainty attaches to this segment of the curve. Nevertheless, the several indicated values are probably approximately correct; and fasting birds at 43 days were found to produce 922 calories per square meter per 24 hours, while the rate of heat production found in fasting birds at $5\frac{1}{2}$ months was only 632 calories.

In embryos of cattle the anterior pituitary glands were found (10) to contain relatively much more prolactin than do these glands in later life stages. Several items of evidence indicate that an increased production and release of prolactin similarly occurs in the pituitary of the very young pigeon; and, further, that this change is probably causally associated with the early period of rapid bodily growth in this animal (7, 11).

Since the brain shares with the pituitary gland the major regulatory powers of the organism, we may refer here to those chemical changes which have been observed in the pigeon brain during growth (12). Perhaps most of the changes found are related to progressive myelination of nerve fibers with age. With older age the percentage of water decreases; protein decreases; total lipoids, cholesterol, and sulphatids increase; phosphatids and extractives show little or no change, at least after an age of 45 days.

The informed reader will readily recall various categories and items of change during growth which are not included in this brief discussion. He will also realize that change in each of the items figured or described above further projects itself into still other related changes in the organism. Finally, the reader may recall that the several marked changes which lead to the sex differences of adults are attained progressively during growth, or at least during that early part of the life cycle which is covered by our graph.

Seasonal Changes

It is well known, even to the layman, that seasons dominate the life and activities of birds in general. Moult, migration, and reproductive state are linked to stages of the earth's progress round the sun. Longer or shorter days, more or less ultraviolet light, more or less heat—these and their concomitants reach into and change vital organs and processes of the pigeon. Some of the observed changes, as indicated by the tops of a series of polygons, are sketched in figure 2. The six items figured were selected because of a general or specific relationship they bear to oxidation rate, or basal metabolic rate, in the organism; and because basal metabolic rate is often considered one of the more constant features of the warm-blooded animal. For all these items our measurements disclose a tidal flux, high and low, running through the annual cycle.

The weight of thyroid glands of ring doves, and likewise of the two other kinds of pigeons examined, undergo the type of fluctuations figured here (13). These birds were kept on the same diet throughout the year, and it is fairly clear that the largest glands (autumn, winter) were also those which most actively produced the thyroid hormone, while the smallest glands (summer) were those that produced least of the metabolism-increasing hormone.

Under certain conditions it was found that the sex ratio in doves (and in certain crosses of doves) is notably influenced by or associated with season (14, 15). It is notable that the highest proportion of males arise from eggs produced in winter and autumn and that this is the period when thyroids are largest and most active. The highest proportion of females arise in summer, when thyroid size and activity are at their lowest.

The polygons which describe the amount of energy stored at different seasons in the egg yolks of these same ring doves tell a parallel story. This storage is least in the eggs of winter, that is, in those which produce an excess of males; it is greatest in the eggs of summer, that is, in those which produce an excess of females (14, 15). For present purposes we need not further consider the way or ways in which season effects changes in the stored energy of eggs and, in a parallel way, affects the sex ratio.

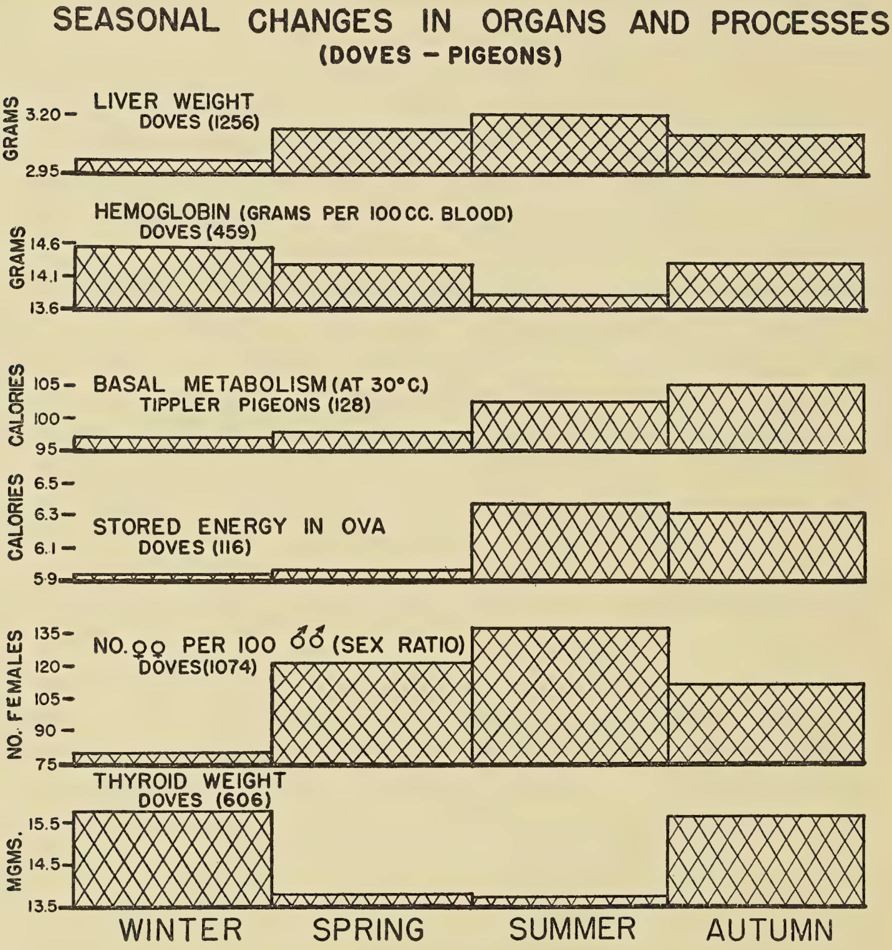


Fig. 2. Seasonal changes in organs and processes (doves - pigeons).

Adequate data on seasonal changes in basal metabolic rate is not now available for any race of ring doves. Indeed, for our present purpose we are forced to utilize data obtained on a somewhat aberrant race of pigeons (2). Tippler pigeons are essentially a dwarf race and our further studies may demonstrate that the seasonal changes in heat production which it shows are atypical of doves and pigeons in general. The high values obtained in autumn, when these animals were subjected to air of greatest cooling power, will probably prove typical; but it seems improbable that in other races the summer values will be found higher than winter values. The more important consideration for present purposes is the obvious fact that basal metabolic rate (as measured at 30° C.) fluctuates with the season. If we should base calculations upon measurements made at 20° C. on this same race we would find that this seasonal fluctuation (spring to autumn) attains 15 per cent in males and 8 per cent in females.

Hemoglobin carries to the tissues the oxygen used in heat production. In ring doves the quantity of hemoglobin varies with season; it is 5.2 per cent more abundant in winter than in summer (16). In parallel measurements made on pigeons this difference was 3.2 per cent. It may be added that in this same study the number of red blood cells, as counted in 1583 doves and pigeons, shows a seasonal change which is wholly comparable to the changes observed in hemoglobin. These cells were most numerous during autumn and winter; they were fewest in summer—fewer then by 10 per cent than in autumn.

The liver is the seat of many and vitally important metabolic processes. Its size fluctuates with the seasons (17). Reference to the graph will show that its weight changes at each season are quite the opposite of simultaneous weight changes in the thyroid gland. Though this is not given a place on the graph, the spleen shows seasonal fluctuations wholly comparable with those observed in the liver. Again, testes and ovaries show seasonal weight changes (18) which are the opposite of those observed in thyroids but quite parallel with those observed in the liver.

Also omitted from the graph is the notable fact that the calcium of the blood undergoes change with the seasons. It is lowest in winter and highest in summer (19). The higher values probably reflect an increased activity of the parathyroid glands. Moreover, these higher values are associated with the season of larger and more active gonads; and female sex hormones, produced by the more active ovaries of spring and summer, have great power (through the parathyroids) to increase the blood calcium (20). Others have shown that the amount of comb growth obtainable from a given quantity of male hormone (androsterone) is 100 per cent greater in winter than in summer (21).

Seasonal tides run broad and deep within the organism. They affect major aspects of the organism's performance, structure, and response. They are especially forceful in connection with endocrine-controlled organs and processes.

Changes in the Reproductive Cycle

Apart from the winter season (if and when fully exposed to cold) the life of an adult pigeon is a succession of reproductive cycles.

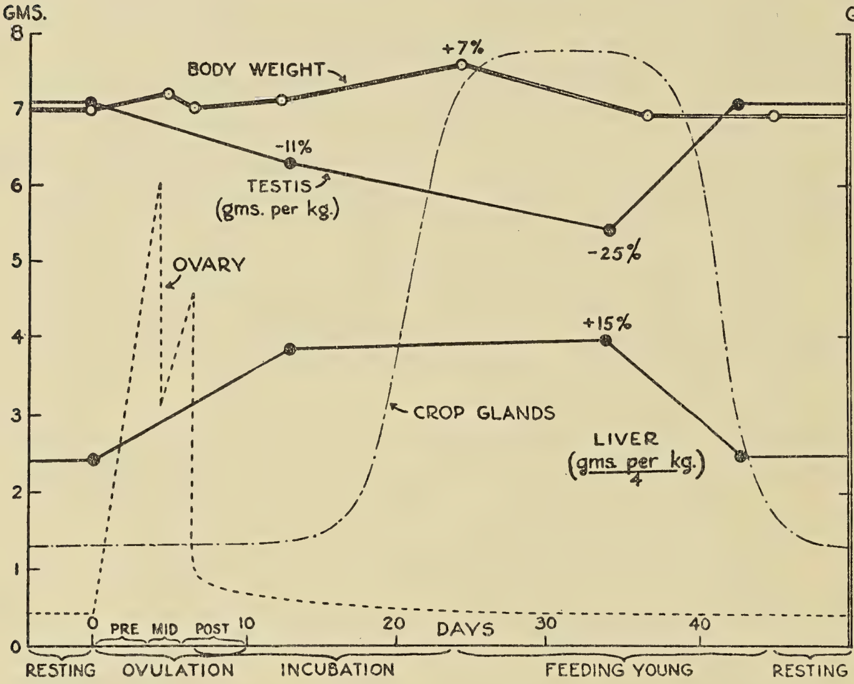


Fig. 3. Body and organ-weight changes during reproductive cycle of pigeons.

Each such cycle is attended by extensive somatic and behavioristic change (22, 23, 7). Certain of these changes (exclusive of changes in females at ovulation) are indicated on figure 3.

At one point in a 50- to 60-day cycle the body weight of the animal will have increased by about 7 per cent and at another point it will have decreased by a similar amount (1). In both males and females an increased weight accompanies the final days of nesting (or incubation) of the eggs; only in small part, however, is the increased body weight directly ascribable to the much enlarged crop glands of this period. The maximum decrease in weight coincides with the termination of feeding of young.

The most marked structural change of the cycle (apart from its ovulation phase) is found in the crop glands. Toward the end of incubation of the eggs the two lateral walls of the crop become 5 to 15 times thicker than usual and produce large amounts of "crop milk." With the cessation of feeding of young these parts of the crop wall cease their activity and become paper thin.

Some reduction of size or weight of the testes has been observed to accompany the incubation of eggs and subsequent feeding of young (7). One should bear in mind that this reduction occurs while crop glands are enlarged as a direct result of the output of prolactin from the bird's pituitary gland (24, 25); and related to this is the further fact that adequate amounts of prolactin cause both testes and ovaries of birds to atrophy as the result of prolactin's ability to stop the output of follicle-stimulating hormone from the pituitary gland (26).

The liver is observed to undergo a 15 per cent increase in weight during the cycle, and this increase tends to coincide with the period of weight decrease in the testes (7). In this case too it appears that prolactin, the hormone that so greatly stimulates the crop glands, also induces enlargement of the liver (27, 11).

Not included in the graph but of much significance in this account are the changes in behavior associated with the reproductive cycle (22). It is fairly clear that the output of sex hormones from the unrepressed testes and ovaries of the "resting" and "ovulation" stages of the cycle are causally related to the persistent courting and frequent copulations which characterize those periods. Again, the broodiness, or incubation behavior, which characterizes the period during which eggs are brooded is found to be causally related to the release of prolactin (28, 29). One may hope to witness few more striking changes in an animal than those involved in the passage of a pigeon from the copulatory to the broody phase; yet the time interval between these two phases may be no more than 48 hours.

Changes at Ovulation

The blood and many important organs of the female pigeon undergo rapid and extensive changes in connection with the dehiscence of a pair of ova from the ovary (30). Several of these changes are indicated on the curves of figure 4. A short statement concerning each of these items of change may be helpful.

The changes pictured on the graph cover a period of about 11 days, time being read on the base line from left to right. It will be of general aid to an understanding of this graph if we first examine the curves for the two ova (first ovum, second ovum), which detach themselves from the ovary at an interval of 44 hours. The "first ovum" begins an accelerated rate of growth at about 108

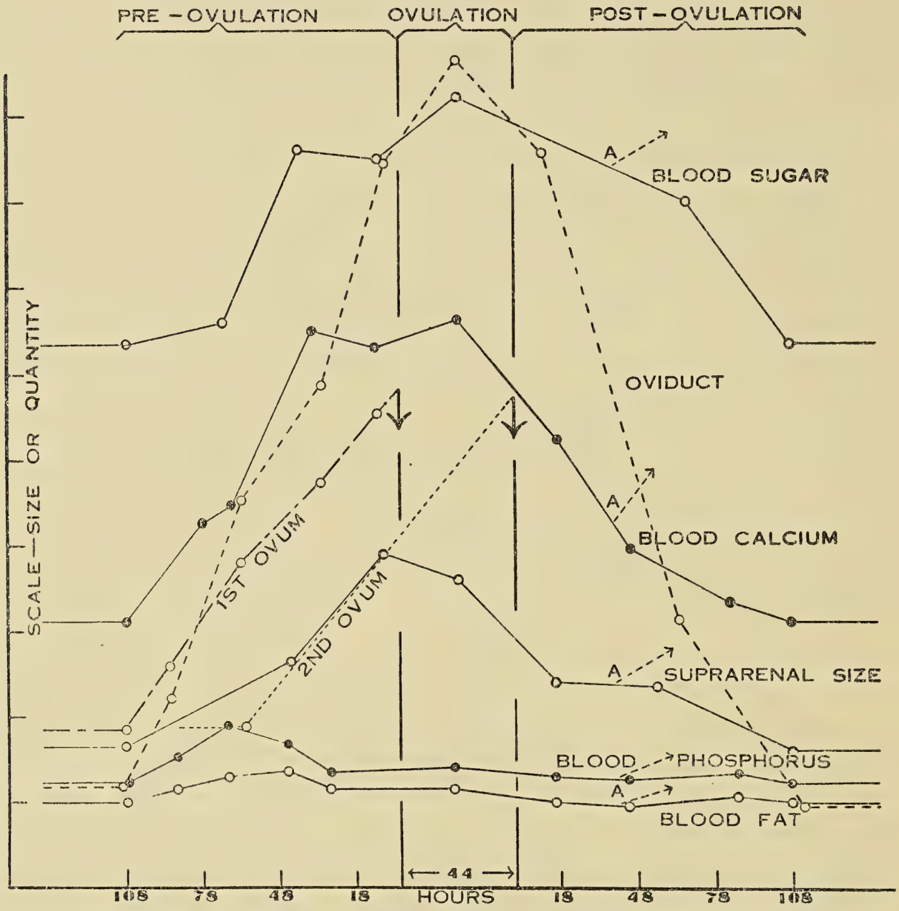


Fig. 4. Metabolic changes at ovulation.

hours before it breaks away from the ovary. This increased rate of growth occurs suddenly, and the succeeding rate of growth is about 25 times the former rate (30). The entire series of events pictured in the graph begins within a few hours of the time that this "first ovum" struck up its new rate of growth. This acceleration of growth rate in the ovum is now known to be induced by a nearly coincident change in the anterior pituitary gland, a change involving an increased release of follicle-stimulating hormone to the blood.

The oviduct (uterus) begins to enlarge with the first signs of increased growth rate in the ovum. Less than 130 hours thereafter the weight of this organ has increased by about 1000 per cent. This phenomenal growth is caused by increased release of female sex hormone from the newly stimulated ovary (31, and unpublished data).

The suprarenal glands undergo considerable enlargement at this period (32). The most extensive enlargement found in any race of pigeons is figured in the graph, but in four types of doves and pigeons an average enlargement of nearly 40 per cent was found. Certainly the cortex of the adrenal shares in this hypertrophy; it is not certain that the adrenal medulla is involved in this change. The causal basis of this enlargement is not definitely known, though the anterior pituitary (perhaps also the ovary) is doubtless involved.

The blood sugar of ring doves seems, as figured on the graph, to undergo increase at the ovulation period (33), though later tests (unpublished) on pigeons have shown smaller and less consistent increases of the blood sugar at this period. It now seems probable that this glycemia results from an increased production of cortical hormone at this period (34).

Blood fat and lipoid phosphorus of the blood may be considered together. These substances increase rapidly in the early pre-ovulation period and become practically normal at the actual time of ovulation (35). These increases reach 35 per cent and 50 per cent, respectively. In fowl, where during egg production ova in the early pre-ovulation phase are constantly present, both blood fat and lipoid phosphorus are continuously present in nearly double their normal amounts (36). The cause of these pronounced changes in the blood is still obscure, but it is probably to be sought in a change in output of one or another hormone of the anterior pituitary and ovary.

The calcium of the blood is doubled within a period of 4 or 5 days (19). The parathyroid hormone apparently mobilizes this additional calcium from the bones; however, the parathyroids, directly or indirectly, are in turn stimulated to this greater activity by a coincident increase of female sex hormone at this period (20). In this connection it is notable that others have shown that a pronounced cyclic ossification of the red marrow of the long bones of the leg occurs in coincidence with this ovulation period (37).

Unrepresented on the graph but important to this subject are the histological and cytological changes observable in the thyroid (38) and anterior pituitary (7) at this period. Again, evidence now published only in part indicates that at ovulation the basal metabolic rate is increased by 5 to 10 per cent (39). Notable too is the fact that the entire activity of one large and important organ, the oviduct, is strictly limited to this ovulation period, and that in a 300-gram pigeon it may then secrete about 20 grams of egg albumen and 2 grams of eggshell within less than 4 days.

Daily Rhythms

Daily rhythms in all higher animals are so obvious and so important that they have for long been taken largely into account in most quantitative biological studies. For example, such irregular rhythms as are connected with the taking of food (secretion, digestion, absorption, peristalsis, etc.) are plainly reflected by changes in blood and tissues. On the other hand, one or two regular rhythms of more elusive nature are notable here.

The body temperature of the pigeon is nearly 1.0° C. lower during 4 hours of the night (1:00–5:00 a. m.) than during another 10 hours of the day. And, concurrently with the hours of lowest body temperature, a very marked lowering of blood pressure probably occurs (40). The effects, direct or indirect, of these daily changes during 4 hours are sufficient to leave a discoverable mark on the growing feathers of most birds of all species (40). A rhythm which is still but little understood determines that the dehiscence of an ovum from the ovary shall occur at a particular early hour of the night (32). Further, the pigeon consistently tends to "lay" her eggs at a definite (daylight) hour, though in this case (unpublished data) the transient prevalence of sunshine or clouds exerts a measurable effect.

Concluding Statement

Outward form and features of a higher organism give impressions of constancy which are proved illusory by measurement of varied, profound, regular and irregular internal change. Much of the constancy of an organism is a constancy of change. Our measurements are made not merely upon an organism but upon a changing organism.

What is the basic source of these many slow and swirling tides of change? This source doubtless shares something of the complexity that attends all life processes. But it is also evident that the several categories of change discussed in this presentation rest largely upon mediations of the anterior pituitary gland. The nervous system and brain seem less involved. Anterior pituitary and brain together direct these numerous flowing adjustments which characterize and propagate the organism.

LITERATURE CITED

1. Riddle, O., and P. F. Braucher, 1934. Body Size Changes in Doves and Pigeons Incident to Stages of the Reproductive Cycle, *Amer. Jour. Physiol.*, vol. 107, pp. 343–347.
2. Riddle, O., G. C. Smith, and F. G. Benedict, 1934. Seasonal and Temperature Factors and Their Determination in Pigeons of Percentage Metabolism Change per Degree of Temperature Change, *Amer. Jour. Physiol.*, vol. 107, pp. 333–343.
3. Riddle, O., 1928. Growth of the Gonads and Bursa Fabricii in Doves and Pigeons, with Data for Body Growth and Age at Maturity, *Amer. Jour. Physiol.*, vol. 86, pp. 248–265.
4. Riddle, O., D. R. Charles, and G. E. Cauthen, 1932. Relative Growth Rates in Large and Small Races of Pigeons, *Proc. Soc. Exper. Biol. and Med.*, vol. 29, pp. 1216–1220.
5. Riddle, O., and P. Frey, 1925. The Growth and Age Involution of the Thymus in Male and Female Pigeons, *Amer. Jour. Physiol.*, vol. 71, pp. 413–429.
6. Riddle, O., and J. P. Schooley, 1935. Absence of Follicle-stimulating Hormone in Pituitaries of Young Pigeons, *Proc. Soc. Exper. Biol. and Med.*, vol. 32, pp. 1610–1614.
7. Schooley, J. P., and O. Riddle, 1938. The Morphological Basis of Pituitary Function in Pigeons, *Amer. Jour. Anat.*, vol. 62, pp. 313–349.

8. Riddle, O., and G. E. Cauthen, 1938. Erythrocyte Number in Young Pigeons and Its Relation to Heredity, Growth and Metabolism, *Amer. Jour. Physiol.*, vol. 122, pp. 480–485.
9. Riddle, O., T. C. Nussmann, and F. G. Benedict, 1932. Metabolism during Growth in a Common Pigeon, *Amer. Jour. Physiol.*, vol. 101, pp. 251–259.
10. Bates, R. W., O. Riddle, and E. L. Lahr, 1935. An Assay of Three Hormones Present in Anterior Pituitaries of Seven Types of Cattle Classified for Age, Sex and Stage of Reproduction, *Amer. Jour. Physiol.*, vol. 113, pp. 259–264.
11. Bates, R. W., O. Riddle, E. L. Lahr, and J. P. Schooley, 1937. Aspects of Splanchnomegaly Associated with the Action of Prolactin, *Amer. Jour. Physiol.*, vol. 119, pp. 603–609.
12. Koch, M., and O. Riddle, 1919. Further Studies on the Chemical Composition of the Brain of Normal and Ataxic Pigeons, *Jour. Comp. Neurol.*, vol. 31, pp. 83–110.
13. Riddle, O., and W. S. Fisher, 1925. Seasonal Variation of Thyroid Size in Pigeons, *Amer. Jour. Physiol.*, vol. 72, pp. 464–487.
14. Riddle, O., 1917. The Control of the Sex Ratio, *Jour. Wash. Acad. Sci.*, vol. 7, pp. 319–356.
15. Riddle, O., 1927. Proofs and Implications of Complete Sex-Transformation in Animals, *Verhandl. 1. Internat. Kong. für Sexualforsch.*, vol. 1, pp. 193–218.
16. Riddle, O., and P. F. Braucher, 1934. Hemoglobin and Erythrocyte Difference According to Sex and Season in Doves and Pigeons, *Amer. Jour. Physiol.*, vol. 108, pp. 554–566.
17. Riddle, O., 1928. Sex and Seasonal Differences in Weight of Liver and Spleen, *Proc. Soc. Exper. Biol. and Med.*, vol. 25, pp. 474–476.
18. Riddle, O., 1925. Reciprocal Size Changes of Gonads and Thyroids in Relation to Season and Ovulation Rate in Pigeons, *Amer. Jour. Physiol.*, vol. 73, pp. 5–16.
19. Riddle, O., and W. H. Reinhart, 1926. Blood Calcium Changes in the Reproductive Cycle, *Amer. Jour. Physiol.*, vol. 76, pp. 660–676.
20. Riddle, O., and L. B. Dotti, 1936. Blood Calcium in Relation to Anterior Pituitary and Sex Hormones, *Science*, vol. 84, pp. 557–559.
21. David, K., 1938. Über den Einfluss der Saison auf die Empfindlichkeit des Hahnenkammes, *Acta Brevia Neerlandica*, vol. 8, pp. 133–136.
22. Whitman, C. O., 1919. Posthumous Works, vol. III: The Behavior of Pigeons, Carnegie Inst. Wash. Pub. No. 257.
23. Riddle, O., 1927. On Endocrines and Organisms, *Amer. Naturalist*, vol. 61, pp. 481–502.
24. Riddle, O., and P. F. Braucher, 1931. Control of the Special Secretion of the Crop-Gland in Pigeons by an Anterior Pituitary Hormone, *Amer. Jour. Physiol.*, vol. 97, pp. 617–625.
25. Riddle, O., R. W. Bates, and S. W. Dykshorn, 1933. The Preparation, Identification and Assay of Prolactin—a Hormone of the Anterior Pituitary, *Amer. Jour. Physiol.*, vol. 105, pp. 191–216.

26. Bates, R. W., O. Riddle, and E. L. Lahr, 1937. The Mechanism of the Anti-gonad Action of Prolactin in Adult Pigeons, *Amer. Jour. Physiol.*, vol. 119, pp. 610–614.
27. Riddle, O., 1933. Differentiating Some Functions of the Anterior Pituitary Hormones, *Ann. Int. Med.*, vol. 7, pp. 1–11.
28. Riddle, O., R. W. Bates, and E. L. Lahr, 1935. Prolactin Induces Broodiness in Fowls, *Amer. Jour. Physiol.*, vol. 111, pp. 352–360.
29. Lahr, E. L., and O. Riddle, 1938. Proliferation of Crop-Sac Epithelium in Incubating and in Prolactin-injected Pigeons Studied with the Colchicine Technique, *Amer. Jour. Physiol.*, vol. 123, pp. 614–619.
30. Riddle, O., 1927. Metabolic Changes in the Body of Female Pigeons at Ovulation, *Proc. Amer. Philos. Soc.*, vol. 66, pp. 497–509.
31. Riddle, O., and M. Tange, 1926. Some Limitations of the Action of the So-called Follicular Hormone in Birds, *Proc. Soc. Exper. Biol. and Med.*, vol. 23, pp. 648–652.
32. Riddle, O., 1923. Suprarenal Hypertrophy Coincident with Ovulation, *Amer. Jour. Physiol.*, vol. 66, pp. 322–339.
33. Riddle, O., and H. E. Honeywell, 1923. Increased Blood Sugar Coincident with Ovulation in Various Kinds of Pigeons, *Amer. Jour. Physiol.*, vol. 66, pp. 340–348.
34. Riddle, O., 1937. On Carbohydrate Metabolism in Pigeons, *Cold Spring Harbor Symposia on Quant. Biol.*, vol. 5, pp. 362–374.
35. Riddle, O., and F. H. Burns, 1927. Blood Fat and Phosphorus in the Sexes and Their Variations in the Reproductive Cycle, *Amer. Jour. Physiol.*, vol. 81, pp. 711–724.
36. Lawrence, J. V., and O. Riddle, 1916. Sexual Differences in the Fat and Phosphorus Content of the Blood of Fowls, *Amer. Jour. Physiol.*, vol. 41, pp. 430–437.
37. Kyes, P., and T. S. Potter, 1934. Physiological Marrow Ossification in Female Pigeons, *Anat. Rec.*, vol. 60, pp. 377–379.
38. Riddle, O., 1929. Endocrine Regulation of Reproduction, *Endocrinology*, vol. 13, pp. 311–319.
39. Riddle, O., G. C. Smith, and F. G. Benedict, 1933. Basal Metabolism and the Temperature Factor in Brooding Ring Doves, *Amer. Jour. Physiol.*, vol. 105, pp. 428–433.
40. Riddle, O., 1908. The Genesis of Fault-Bars in Feathers and the Cause of Alternation of Light and Dark Fundamental Bars, *Biol. Bull.*, vol. 14, pp. 328–370.

OBSERVATIONS ON EVOLUTIONARY CHANGES IN THE CHROMOSOMES OF SCIARA (DIPTERA)

C. W. METZ

Department of Embryology
The Carnegie Institution of Washington

Dobzhansky has performed a valuable service in his recent treatise "Genetics and the Origin of Species" (Dobzhansky, 1937) by directing renewed attention to the problems of evolution as investigated by current genetic methods, centering particularly in the study of the chromosomes. As he points out, there is strong reason for believing that the basis of most evolutionary change is chromosome change, and that "mutation" plays a primary rôle. Many investigators are contributing to the subject, directly or indirectly, at the present time through genetic and cytological studies. And it is noteworthy that for the most part the main advances come where cytological observation and genetic analysis go hand in hand.

In the present paper brief consideration is given to some of the evolutionary problems presented by conditions in the fungus flies, *Sciara*, which have been under observation in this laboratory for several years. As will be evident from the papers cited in the following pages, the material presented here has come from the cooperative investigations of several workers, using the methods of cytology, genetics, and embryology. That dealing with the effects of irradiation has come as a result of the cordial cooperation of the X-ray Staff of the Fixed Nitrogen Research Laboratory of the United States Department of Agriculture, in Washington, and the Radiation Staff of the Kelly Hospital in Baltimore, to whom we express especial appreciation.

Although our primary efforts, up to the present time, have necessarily been devoted to straight cytological, embryological, and genetic analysis, in order to get some understanding of the peculiar phenomena of reproduction and inheritance in *Sciara*, it has long been evident that one of the most interesting aspects of the study should eventually center in the mode of evolution of the bizarre genetic mechanism found here, and in other evolutionary problems presented by these flies. Additional impetus has recently been given to this latter aspect of the work by the discovery of two closely related species which will hybridize and thus permit a more intimate study of species relationships (cf. Metz and Lawrence, 1936, 1938). Likewise, it seems probable from present evidence that a still more refined analysis of certain features may be made by studying racial or varietal relationships here. In certain important respects the phenomena coming in this latter category, like most others thus far dealt with in *Sciara*, seem to differ from those found in the better-known representatives of other groups, such as *Drosophila*—a fact giving them added interest.

The present account is devoted primarily to one restricted topic, dealing with the significance of small chromosomal differences between closely related individuals or species; but in the latter part of the paper attention will be given briefly to a few grosser phenomena and more general problems which are of interest in the same connection. It should be emphasized at the outset that, in the present limited state of knowledge regarding the present field, attention must necessarily be devoted largely to search for significant facts and to establishing probabilities. Few definite conclusions of general nature can as yet be reached. Another point which should perhaps be emphasized is that where conditions are described here as present in *Sciara* it is understood that the statements refer only to the particular one or more species under consideration. It is possible that, particularly as regards mutation and chromosome changes, conditions may differ materially in different species.

MUTABILITY OR MUTATION RATE IN SCIARA

If mutation or rate of mutation is an important factor in evolution, it would seem possible to throw light both on its nature and on the rôle it plays, by comparative study of conditions in different groups of organisms. Several groups of animals and plants are known in which the mutation rate is definitely high, e.g., *Drosophila* (fruit flies) and maize (corn). In *Sciara*, on the other hand, mutation, as judged by the appearance of visible mutant characters or of lethals, seems comparatively to be a very rare phenomenon. Comparison of conditions in *Sciara* with those in *Drosophila*, therefore, would seem to provide a promising mode of attack on the problem just stated. The two genera belong to the same family, which makes it possible to use comparable methods of investigation, and one (*Drosophila*) belongs among the "higher" Diptera while the other belongs among the "lower."

In comparing rate of mutation in the two genera it is not yet possible to state the difference in numerical terms, for no absolute basis of comparison is available. Nevertheless, it seems clear that the difference is great. The writer has worked extensively on *Drosophila* as well as *Sciara*, using comparable methods. Likewise, among those who have cooperated in the work on *Sciara* four have also investigated *Drosophila* in similar manner. Uniformly the rate of mutation has been found exceedingly low in *Sciara* as compared with that in any of the best-known species of *Drosophila*. Only nineteen mutant characters have been found as a total in study of five species of *Sciara* extending over several years. One species, *S. coprophila* Lintner, was studied intensively in this laboratory for five years by Dr. Helen Smith-Stocking alone, and has likewise been studied by Miss Mildred Moses, Miss Ella N. Hoppe, Dr. J. Paul Reynolds, Miss Helen V. Crouse, the writer, and others, mostly for periods ranging over two to five years, all with the same general result: a total of thirteen mutants (see, e.g., Smith-Stocking, 1936; Crouse, 1938). Only a small portion of this work, of course, was devoted exclusively to search for mutants; but throughout the experiments careful watch was kept for them.

Part of the difficulty of securing mutants in *Sciara* is due to peculiarities in its mode of reproduction (see Metz, Moses, and Hoppe, 1926; Metz, 1931, 1938), but these should not interfere seriously with the appearance or detection of sex-linked characters or dominant characters.

The preceding statements have reference primarily to so-called "spontaneous" mutation; but they appear to hold equally well as regards artificial induction of mutations by irradiation with X-rays or radium. Extensive experiments, using a wide range of dosages and both types of irradiation, are included in the work just referred to. These have given almost no visible mutant characters. Not more than three or four seem to have come as a result of the treatment. Judging simply by experience with the two groups of organisms, therefore, it seems safe to say that comparable efforts and methods will result in acquisition of more than ten times as many mutants in *Drosophila* as in *Sciara*, either spontaneous or induced.

Another fact meriting note in this connection is that the cells or tissues of *Sciara* appear to be exceptionally resistant to irradiation, as judged by comparing sterilizing and lethal doses. It has not yet been possible to give this aspect of the work the attention it appears to deserve, and since the results depend in part on temperature, humidity, and other secondary factors, and no simultaneous treatments of other organisms have been made, details need not be elaborated here.

From the evidence just reviewed it seems clear that the cells or chromosomes of *Sciara* possess some general property making them resistant to irradiation, and some property which tends to inhibit ordinary spontaneous mutation. As a tentative hypothesis the two properties may be considered as related or identical, in which case the solution of the problem is to be sought in the makeup and behavior of the chromosomes.

When the known facts about the chromosomes of *Sciara* are examined from this point of view one feature seems especially significant. It concerns the frequency of chromosome rearrangements, as shown by study of the giant salivary gland chromosomes (see below). Numerous observers have shown that in *Drosophila* inversions of segments of chromosomes frequently occur spontaneously and may readily be induced by irradiation (for references see Dobzhansky and Sturtevant, 1938). Likewise, translocations of pieces of chromosomes to other positions or to other chromosomes occurs, although apparently with less frequency than inversions. It may also be noted that inversions are relatively very numerous in another dipteran genus, *Chironomus* (King and Beams, 1934; Bauer, 1935; Poulson and Metz, unpublished). In *Sciara*, on the other hand, no case of inversion has been detected in our extensive cytological study of three species, including irradiated material, and none has been definitely identified in the species hybrids. This does not necessarily indicate that inversions never occur here, but it leaves little doubt that they are rare as compared with *Drosophila* and *Chironomus*. Presumably the same consideration applies to translocations. Only two cases involving ordinary translocation configurations have thus far been detected in *Sciara*, both apparently due to irradiation. Three cases are known, however, of duplications or "repeats" (Bridges, 1935), in which a group of several bands is normally found in two places some distance apart in the same chromosome (Metz, 1937). These presumably arose through translocation. Probably cases representing translocation will be found in the species hybrids in *Sciara* mentioned above, but none has been identified thus far.

In view of the rapidly accumulating evidence that many, if not most, "mutations" in any organism result from chromosome rearrangements (Muller, 1932; Stadler, 1932) it may be postulated that the low frequency of mutation in *Sciara* is due to the low frequency of these rearrangements. If this is true it reveals at least one chromosome characteristic in *Sciara* which tends toward stability of chromosome organization. The precise nature of this characteristic is unknown, as yet, and cannot be inferred from general evidence because the mechanism of chromosome rearrangement is also unknown. Studies designed to throw light on this feature are under way.

It may be pointed out here that general taxonomic position apparently has nothing to do with the condition or relations just discussed. This seems evident from the relative taxonomic positions of the three groups of organisms concerned. *Drosophila* and *Sciara* differ in both taxonomic position and chromosome behavior, but *Chironomus*, which agrees in general with *Drosophila* in chromosome behavior, belongs, like *Sciara*, among the lower Diptera. In this connection it should, perhaps, also be pointed out that the apparently conservative tendency in chromosome behavior in *Sciara* just considered may be counterbalanced by a tendency toward a different kind of chromosome change. This topic will be treated in the next section. Furthermore, it may be observed that, although placed taxonomically among the lower Diptera, *Sciara* is a highly specialized genus in some respects, especially as regards its mode of sex determination, phenomena of spermatogenesis, chromosome behavior during embryonic development, and related aspects. It is not to be assumed that evolution is necessarily going on here at a slower rate than in such genera as *Drosophila*. This question must remain open for the present.

NATURE AND POSSIBLE EVOLUTIONARY SIGNIFICANCE OF MINUTE CHROMOSOME CHANGES IN SCIARA

Recently a great refinement in the method of analyzing chromosome changes and chromosome relationships has been made possible by discovery of the fact that the giant structures in nuclei of the salivary glands of Diptera are true chromosomes (Kostoff, 1930; Heitz and Bauer, 1933; King and Beams, 1933; Painter, 1933). These giant chromosomes are now being used widely in studying species and racial relationships in the Diptera, particularly *Drosophila* (for review of the literature see Dobzhansky, 1937). Not only are they relatively enormous in size, but they exhibit a complex and relatively constant internal pattern reflecting their longitudinal differentiation, and they are intimately associated in pairs, each pair representing a maternal and a paternal homologue fused side by side (plate 1, figure 3). This latter fact makes it possible accurately to compare the makeup of homologous chromosomes from two parents, whether these parents belong to the same strain or species or to different strains or, in favorable cases, even to different species.

In a study which may well serve as a model for subsequent investigations, Dobzhansky and Tan (1936), using hybrids between *Drosophila pseudoobscura* and *D. miranda*, have called attention to two categories of differences visible in these giant chromosomes. One category includes gross differences which are clearly due to chromosome rearrangements (mainly the types noted above: inversions and translocations). The other involves relatively very small modifications, the nature of which is not yet clear. The latter are revealed in the hybrids in the form of small differences involving only one or very few of the transverse chromatic disks or "bands" visible in these chromosomes.

In *Drosophila* differences of this latter type are apparently rare, both in species hybrids and within individual species, while, as noted above, the gross differences due to chromosome rearrangements are relatively common. In *Sciara*, on the other hand, the situation appears to be just reversed: the minute differences are relatively common and the other type rare. It seems possible that this fact may have significance from the evolutionary standpoint. In *Chironomus*, the only other dipteran genus on which evidence is thus far available, conditions so far as known agree with those in *Drosophila.* Chromosome rearrangements (at least inversions) are relatively very common, apparently more so than in *Drosophila.* Little evidence is available regarding minute differences in *Chironomus*; but extensive general observations on the chromosomes of this genus made by Dr. D. F. Poulson and the writer (unpublished) have revealed only one such case. The data strongly suggest, therefore, that a marked difference exists in these respects between *Sciara*, on the one hand, and *Drosophila* and *Chironomus* on the other. This fact, together with the possible significance of the phenomena themselves, serve to focus present attention on the minute differences in *Sciara.*

Although the study is still in its initial stages the two aspects just mentioned will be considered briefly in the following paragraphs.

In the intensive study of hybrids between *Drosophila pseudoobscura* and *D. miranda*, referred to above, Dobzhansky and Tan report at least twenty-five cases of gross rearrangement (inversions, translocations) but only three clear cases of minute differences of the type under consideration. In one of the latter a "rather heavy disk" is present in one species which is not found at the corresponding locus in the other. The other two cases are similar except that the disk is fainter and apparently it is one member of a pair of disks which is involved (see below). Two or three other cases may represent comparable differences, but the nature of these is not yet fully understood.

In an earlier study of hybrids between *Drosophila melanogaster* and *D. simulans*, Pätau (1935) noted two or three cases which may represent minute differences. These apparently were not clear, however, and it is not certain how they should be interpreted. His figures seem to show clearly that for the most part only chromosome rearrangements are involved, not small localized changes.

In interracial crosses in *Drosophila pseudoobscura*, and likewise in random samples of the populations of this species and of *D. miranda*, Dobzhansky (personal communication) has found little evidence of minute differences. To some extent the apparent rarity of these differences in *Drosophila* may be due to the fact that most observers have made no definite search for them; but it seems clear, nevertheless, that they are rare here as compared with their frequency in *Sciara.*

In hybrids between *Sciara ocellaris* and *S. reynoldsi*, corresponding or homologous chromosomes show numerous small differences of the type under discussion. The exact number is not known; indeed it is too large to estimate at present. Several cases have been described in an earlier paper (Metz and Lawrence, 1938) and additional ones are illustrated in plates 1 and 2, figures 1, 2, 4, 9, 10. However these particular differences between the two species originated, it seems safe to assume that they appeared at different times, not through any one sudden reorganization. It would be expected, therefore, that similar differences would appear now and then, singly or in small numbers, in ordinary strains or populations of individual species. This is found to be the case in *Sciara*. In the best-known species, *S. ocellaris* Comst., fifteen such differences have been found in the course of our work—most of them before any search was made for the purpose. Almost certainly numerous other examples are present in this same material, for the study has been confined mainly to a few chromosome regions. A few of the well-known examples have been described previously in connection with other problems (see especially Metz, 1937). In one case, at least, a difference is found here which appears to be identical with one observed in the hybrids (compare figures 2 and 3, plate 1). In some cases the differences have been found in descendants of flies which had been irradiated, but most of them have been found in "wild-type" stocks, descended from individual females taken in nature, without any artificial treatment other than that incident to cultivating the stocks in the laboratory. It seems almost certain that the results from these stocks reflect conditions occurring in nature.

No systematic study has yet been made in *Sciara* of conditions in hybrids between different wild strains within a species. It may be observed, however, that in the case of some of the differences under consideration, individuals of all three expected types have been found in the laboratory stocks examined. That is, in some individuals a chromatic disk is present at a given locus in one chromosome but not in its homologue; in other individuals the disk is present in both homologues; and in others it is not present in either. Thus both types of homozygous individuals, as well as the heterozygous, are found (Metz, 1937). Although perhaps not true for these particular cases, it seems probable that comparable homozygous strains occur in nature which, when crossed, would give the heterozygous condition. No attempt has yet been made to associate any external morphological characters with these chromosome differences. Although some of the differences have been found in "mutant" stocks it is not certain that they are responsible for the mutant characters.

Morphologically the differences under discussion range from those involving only one, apparently single, very small disk (figures 5, 6, plate 1) up to those involving a small cluster of probably three or four disks (figures 3, 7, 8). In general they appear to represent changes of a purely quantitative nature; but one case in the species hybrids, and possibly one or two within a species, have characteristics which suggest that qualitative chemical changes may also be involved, as will be indicated below. Considering only the quantitative aspects, the question immediately arises as to whether the mechanism responsible for these small changes is the same as that giving the gross alterations. If it is, the difference between *Sciara* and *Drosophila* presumably reflects a difference in the operation of a common mechanism. If it is not, however, there may be two mechanisms or processes involved which could be distinguished through further study. No definite answer can yet be given to this question, but a few suggestions may be considered.

Of the fifteen minute differences observed thus far in *Sciara ocellaris*, seven are in chromosome *A*, three in chromosome *B*, one in chromosome *C*, one in the *X* chromosome, and one has not yet been "placed." (The distribution probably merely reflects the fact that *A* and *B* have been studied much more than the others, and that *A* is twice as long as any of the others.) Eight of the differences involve what appears to be, or may be, only a single disk each (figures 5, 6, at *a*). On the other hand, at least four cases are of the type involving what appears to be a very small cluster of bands. It is possible that two different categories are represented here and that they reflect the operation of different processes, but there is no way of deciding the point at present. In any event it is the former type (single band differences) which seems of special interest, because that type is farthest removed from the gross chromosome rearrangements.

Among the eight "single band" differences it is possible that further subdivision should be made. In six cases the disk or band which has no mate opposite it in the other chromosome does have adjacent to it in the same chromosome a similar band. In other words, it is one member of what appears to be a pair of bands, and what is lacking in the other chromosome is similarly just one member. There is still at least one band of this kind present at this general location in each homologue. Pairs of bands are found commonly in these giant chromosomes (Alverdes, 1912) and are presumed to have arisen through duplication of single bands. Cases of the present kind, therefore, presumably represent such duplication, or else the loss of a member of a pair following a previous duplication. The other two cases, however, appear in the light of present evidence to be of a different type. Unfortunately, these cases are difficult to study because of local chromosome disturbance presumably caused by the deficiency itself; but they at least suggest a situation in which a single band, without any adjacent mate, in one chromosome is not matched by anything at the corresponding locus in the other. Apparently similar examples are also observed in the species hybrids (figure 4). And one case of apparently the same kind is found in the *Drosophila* hybrids, as noted above.

If the above interpretation is correct as to the nature of the two types of single band differences it is possible that they differ in mode of origin. The former type appears to involve duplication, at least as the first step. Although this could occur by translocation (Muller, 1935) or inversion, as do presumably the large "repeats" mentioned above, it seems probable that another mechanism is involved. Division of the gene or locus in a transverse, rather than the usual longitudinal, plane is one possibility; "unequal crossing over" (segmental interchange which is not exactly reciprocal) is another. There is little choice between the latter two mechanisms at present, but one line of evidence seems to favor unequal crossing over. If transverse division (longitudinal doubling) were responsible it would seem probable that three, four, or more duplicates might often result, which does not appear to be the case. Although effective in doubling a disk, unequal crossing over, by its very nature, would presumably not have the same tendency toward accumulation of duplicates. The problem is complicated, however, by the fact that multiplication of individual loci might have marked detrimental effects on the viability of the flies and hence prevent accumulation by any means.

The second type of single band difference may be accounted for by assuming either the complete loss of a single band which is not itself a member of a pair, or the translocation of a single band to another position. The former change would automatically accompany the type of duplication just considered, for the process of unequal crossing over which inserted an additional band in the recipient chromosome would remove that band from the homologous donor chromosome. If the loss did not have a deleterious effect in heterozygous condition it would result in the situation observed, and if complete loss could be tolerated some descendants of such flies should lack this particular band or locus entirely. It would be very difficult to demonstrate the occurrence of the latter process, however, because of the alternative possibility, namely, that the locus or disk which appears to be absent is merely shifted to some other position. But it should be observed that, as an alternative explanation, shift in position is not very different from duplication. It would presumably lead to the production of the same three classes of individuals as duplication would, namely, those possessing the disk in a homozygous balanced condition, those heterozygous for it, and those lacking it entirely. The reason for this is that, if the apparently missing disk were in another region of the same chromosome, crossing over should sometimes remove it from this chromosome and add it to the homologue which already possesses one at the locus under observation. This would leave the recipient chromosome with two disks and the donor chromosome with none. Similarly, if the shifted disk were in a nonhomologous chromosome random segregation at gametogenesis should lead to corresponding conditions as regards the presence or absence of the disk in individual flies.

On either hypothesis, therefore, these single band differences seem to indicate the presence of a mechanism which may serve either to increase or to decrease the disks individually, and thus modify the sum total of germinal material. (The material between the disks appears to go along with the individual disks in any alterations of this kind; Metz, 1937). So far as duplication of individual disks is concerned the situation seems simple, both here and in the case of the paired disks; but as regards loss of disks the significance is uncertain. Loss of a particular disk under observation seems to be required, whether or not the condition is perpetuated, but this, of course, does not necessarily mean that that particular kind of material is entirely lost from the germ plasm. It is always possible that elsewhere in the chromosome group replicas of this disk exist in balanced condition—one in each homologue—so that this particular genic material will continue to be present in "double dose" regardless of what happens to the disk under observation. It seems entirely possible that individual genic materials may be repeated many times throughout the chromosome group of a higher organism, particularly in view of the now well-established fact that the effect of a gene may differ greatly according to its position (the so-called "position effect").

Although less attention has been given to the small differences involving more than one disk, the same general considerations seem to apply to them. In some cases disks in one homologue appear to be entirely lacking in the other and to have no adjacent duplicates. One such is shown in an earlier paper (Metz, 1937, figs. 13–15); another is shown in plate 2, figure 7; and one which may be of this type is shown in plate 2, figure 8.

As already suggested, it is too early to gauge the significance or importance of the findings reviewed above. The evidence indicates that internal chromosome changes in *Sciara* (at least *S. ocellaris*) differ markedly from those found in *Drosophila*, and that there is likewise a marked difference in rate of what may be called "ordinary" mutation in the two genera. But the meaning of these differences from the evolutionary standpoint is not yet clear. Similarly, little appears to be gained at present by attempting to determine whether the gross alterations or the minute ones should be the more effective in bringing about evolutionary changes. Superficially it would appear that more drastic evolutionary changes might result from the former type and smaller or more gradual ones from the latter; but even this is by no means certain. Neither is it certain which would provide better material for actual evolutionary change if the difference does exist.

Present evidence reveals no obvious reason why minute chromosome changes should not occur frequently in *Drosophila* or gross changes occur frequently in *Sciara*. Presumably, as already stated, the chromosomes differ in constitution or behavior in the two genera. A possible line of approach to this problem has been suggested in an earlier paper (Metz, 1934), where the hypothesis was advanced that chromosome rearrangements, etc., are dependent upon viscosity characteristics of the chromosome matrix or "sheath." If this is correct a difference of that nature between the chromosomes of *Drosophila* and *Sciara* might be revealed by further study. Such study has been begun.

Possible Localized Qualitative Changes in Sciara Chromosomes

In the preceding section attention was necessarily devoted primarily to changes of a quantitative nature. Indeed, so far as the present subject is concerned, practically all the evidence thus far obtained from salivary gland chromosomes, regardless of source, has involved quantitative rather than qualitative characteristics. The reason for this, of course, is not that qualitative differences are necessarily absent, but that they are difficult to detect through morphological study. The quantitative modifications already observed in the chromosomes of several organisms (see review in Dobzhansky, 1937) provide a means of satisfying one of the well-known requirements of evolutionary change—addition or multiplication of germinal materials—but they do not satisfy the complementary requirement of qualitative changes in these materials after the increase in amount has been effected. There is general agreement that such changes must occur, but practically no direct observational evidence of them.

It has long been thought that "mutations" provide indirect evidence of such changes; but with the present uncertainty surrounding the phenomena of mutation, due in part to the evidence of "position effects" noted above and in part to the present confusion as to what a "gene" is in terms of actual chromosome materials, the significance of the genetic data has become uncertain. Some supposed "gene mutations" are now known to represent gross chromosome rearrangements. Others represent smaller rearrangements or losses or additions, and still others involve no visible alterations. The genetic concept "gene," therefore, appears to apply to compounds of widely different sizes, grading down to the submicroscopic. At present there seems to be no obvious method of distinguishing genetically between quantitative and qualitative mutational changes. This fact serves to give special interest to any cytological evidence of localized qualitative changes in chromosomes. Partly on that account we have in this laboratory been devoting special study to the qualitative characteristics of the salivary gland chromosomes. Generalized qualitative changes, of apparently reversible nature, have been discussed in earlier papers (Metz, 1935*a*, 1935*b*; Metz and Lawrence, 1937; Doyle and Metz, 1935; Poulson and Metz, 1938). Likewise, it has been noted that, superimposed upon the general pattern of transverse chromatic disks separated by narrow achromatic spaces in these chromosomes, there are more delicate localized patterns which presumably reflect qualitative differences in the materials in different regions (Metz, 1935*b*; Metz and Lawrence, 1937). These localized patterns, like the visible structural characteristics of the chromosome as a whole, may normally undergo considerable change, presumably as a result of changes in physiological conditions in the nucleus. Great caution must be observed, therefore, in attempting to identify cytologically any localized qualitative changes of a permanent (mutational) nature. Yet this method seems as reliable as any available at present.

Two cases which suggest localized qualitative change of this kind have been observed in *Sciara*. One is found in the species hybrids noted above and has been described briefly before (Metz and Lawrence, 1936, 1938). It involves structural differences in what appear to be homologous loci in the two species, indicated by arrows in plate 2, figures 9, 10 (see also Metz and Lawrence, 1938, fig. 11). In *S. ocellaris* the region, which is near end 1 of chromosome *B*, is usually coarsely vesiculated and often shows large, block-like chromatic granules (Metz, 1935*b*, figs. 4A, 5A, B, C; Poulson and Metz, 1938, figs. 16B and 17A). Sometimes, however, it is sharply "banded," with the disks distinct (Metz, 1935*b*, fig. 4B; Poulson and Metz, 1938, fig. 17B). In *S. reynoldsi* the apparently homologous region has always been found to exhibit the sharply banded condition (Metz and Lawrence, 1938, fig. 11B). Alone, this evidence might not be considered important; but the data from hybrids between the two species add the necessary touch. Here the two chromosomes are present in the same cell and subject to the same influences, yet each retains its own characteristics.

That from *ocellaris* shows the coarse vesiculation and that from *reynoldsi* the sharp bands at the location in question (plate 2, figures 9, 10; also Metz and Lawrence, 1938, fig. 11A). The question arises, of course, as to whether or not there is actual homology of parts in this case. It is difficult to prove a homology, but the evidence points strongly toward its existence. There is close correspondence in general pattern throughout this region in the two chromosomes and the homologues sometimes show paired association in the hybrids (figure 9). No fusion has been observed at the particular loci in question, but if a qualitative difference exists here, as postulated, fusion would not be expected.

Although the data just reviewed are not conclusive, they represent the best cytological evidence, known to the writer, of localized qualitative change within a chromosome. If such a chemical change is actually involved it seems reasonable to assume that it may have evolutionary significance.

The other case, or type of cases, has very doubtful significance, and is only considered because of the scarcity of evidence on the subject. It involves slight visible differences in apparently homologous bands in ordinary specimens within a species (i.e., not species hybrids) as well as in hybrids. One example, illustrated in an earlier paper (Metz, 1937, locus *b* in figs. 7 and 8*a*) is shown at *a* in plate 1, figures 5 and 6. A deficiency is involved here. There are two similar bands in one homologue and just one in its mate. This one "band" is often in the form of a row of rather widely spaced granules, in contrast to the more nearly continuous nature of the two bands in the other chromosome. Probably it is the presence of the deficiency here which is responsible for the difference in appearance; but since other similar cases are known, and since the nature of minute deficiencies is still imperfectly understood, the possibility of qualitative change should not be overlooked.

Evolutionary Changes in the Large "Limited" Chromosomes in Sciara

The present section and the one following deal with an aspect of our general problem in which the mode of approach is essentially the opposite of that just considered. Instead of examining the germ-plasm for evidence of visible changes in it, and then reasoning as to the possible evolutionary effects of these changes, attention will here be given to the effects themselves, and an attempt made to deduce from them the nature of the germinal changes through which they arose. They involve two specific examples reflecting evolutionary change. One concerns the "loss" of a particular chromosome; the other concerns the modification of a genetic process affecting the sex ratios. Attempts to throw light on the mode of evolution of the bizarre "genetic mechanism" in *Sciara* have thus far not been very fruitful because the general "mechanism" appears to be essentially constant in most respects throughout all the fourteen species studied. In the two respects considered here, however, conditions are not uniform.

All but two of the species of *Sciara* thus far studied possess in cells of the germ-line a peculiar, large chromosome referred to as the "limited" chromosome because of this limitation to the germ-line. It is regularly eliminated from somatic cells during very early embryonic or cleavage stages. This chromosome, which may be present singly or in duplicate, and which readily fragments, appears to be relatively "empty." That is, it appears to contain relatively few genes, or a relatively small amount of "vital" material. It is presumably in process of disappearing (Metz, 1931, 1938). Its constant presence and large size in species possessing it, however, leave no doubt that in these it performs a vital function. In two species (*S. ocellaris* and *S. reynoldsi*) it is not found. Superficially the chromosome groups of these two species do not appear to differ from those of several related species except in the absence of this large chromosome. It seems reasonable to assume, therefore, that the vital parts of this chromosome have been transferred to one or more of the other four chromosomes without any obvious change in size or appearance of the latter.

This situation, like that seen in other cases of "empty" chromosomes, presents a series of problems about which relatively little is known as yet (cf. Muller and Gershenson, 1935). Only two points can be treated here. In most other known cases, vital but relatively "empty" chromosomes are sex chromosomes (Y-chromosomes), although Kikkawa (1937) has recently described one which apparently is not. The "limited" chromosome in *Sciara* is not a sex chromosome. In spermatogenesis, however, it remains condensed during the "growth period" in a manner characteristic of sex chromosomes in many other Diptera, suggesting the possibility that it may be the product of a sex chromosome which has now lost its sex-determining genes to the other chromosomes. Another possibility, of course, is that its behavior is due to its relative "emptiness." Further comparative study may throw light on this subject.

Of greater interest is the question of how the "limited" chromosome disappeared. The fact that all but two of the species possess it seems to leave no question that it has disappeared in the case of these two, rather than just made its appearance in the case of the others. Furthermore, the fact that it is full-sized in all the twelve species possessing it, and completely gone (i.e., not merely reduced in size) in the other two, tends strongly to indicate that its actual loss, quantitatively, was a sudden, one-step process. Assuming that its vital parts are still present (in the other chromosomes, which, as noted, do not appear appreciably changed in size), the possibility may be suggested that genes which "occupied" relatively large portions of the "limited" chromosome now take up relatively little space in their new locations, owing to a "position effect."

Evolutionary Changes in the Sex-determining Mechanism

Another unstable or non-uniform condition in *Sciara* is manifested by the different types of sex ratios found in the genus (for review see Metz, 1931, 1938). In some species pair matings regularly give "unisexual" families, that is, families in which the individuals are all or nearly all of one sex, either males or females; in other species they regularly give bisexual families; and in still other species both types of reproduction are found. The mechanism responsible for the production of "unisexual" families has been analyzed in considerable detail (see references just cited) and that involved in production of bisexual families is under investigation at present, especially by Mr. C. B. Davidheiser. As already noted, the former mechanism deviates widely from that found in "ordinary" organisms. It might be expected, therefore, that the mechanism giving rise to bisexual families in *Sciara* would represent a significant intermediate step connecting this mechanism with the "typical" one found in "ordinary" organisms, where the families are regularly bisexual. Possibly an intermediate step is thus represented; but it does not appear to illuminate this particular problem. Rather, the data point to a relatively slight difference between the modes of production of "unisexual" and bisexual progenies. For present purposes this is the more valuable situation, for the small differences or modifications lend themselves more readily to analysis than large ones would. It seems certain that one of these modes of reproduction arose from the other, although there is no certainty as to which way the process has gone. The most significant fact thus far revealed is that the two conditions apparently represent a difference of only a few genes, together perhaps with a small general difference in chromosome behavior which influences the rate of segmental interchange (crossing over) between homologous sex chromosomes. Presumably the difference has arisen by mutation and the resulting modification has become fixed in some species but not in others. Studies now under way are designed to determine the precise nature of the genetic differences involved here and also indicate more specifically their mode of origin.

LITERATURE CITED

ALVERDES, F., 1912. Die Kerne in den Speicheldrüsen der Chironomus-Larve, *Arch. f. Zellforsch.*, vol. 9, pp. 168–204.

BAUER, HANS, 1935. Der Aufbau der Chromosomen aus den Speicheldrüsen von Chironomus Thummi Kiefer, *Ztschr. f. Zellforsch.*, vol. 23, pp. 280–313.

BRIDGES, C. B., 1935. Salivary Chromosome Maps, *Jour. Heredity*, vol. 26, pp. 60–64.

CROUSE, H. V., and HELEN SMITH-STOCKING, 1938. New Mutants in Sciara and Their Genetic Behavior, *Genetics*, vol. 23, pp. 275–282.

DOBZHANSKY, TH., 1937. Genetics and the Origin of Species. 364 pp. New York, Columbia Univ. Press.

——— and A. H. STURTEVANT, 1938. Inversions in the Chromosomes of Drosophila pseudoobscura, *Genetics*, vol. 23, pp. 28–64.

——— and C. C. TAN, 1936. Studies on Hybrid Sterility, III, *Ztschr. f. ind. Abs. u. Vererb.*, vol. 72, pp. 88–114.

DOYLE, W. L., and C. W. METZ, 1935. Structure of the Chromosomes in the Salivary Gland Cells in Sciara (Diptera), *Biol. Bull.*, vol. 69, pp. 126–135.

HEITZ, EMIL, and HANS BAUER, 1933. Beweise für die Chromosomennatur der Kernschleifen in den Knäuelkernen von Bibio hortulanus L. (Cytologische Untersuchungen an Dipteren, I), *Ztschr. f. Zellforsch. u. Mikr. Anat.*, vol. 17, pp. 67–82.

KIKKAWA, H., 1937. The Inert Chromosomes of Drosophila ananassæ, *Cytologia*, Fujii Jubilee Vol., pp. 125–128.

KING, R. L., and H. W. BEAMS, 1933. Somatic Synapsis in Chironomus, *Anat. Rec.*, suppl., vol. 57, p. 89.

——— 1934. Somatic Synapsis in Chironomus with Special Reference to the Individuality of the Chromosomes, *Jour. Morphol.*, vol. 56, pp. 577–592.

KOSTOFF, D., 1930. Discoid Structure of the Spireme, *Jour. Heredity*, vol. 21, pp. 323–324.

Metz, C. W., 1931. Unisexual Progenies and Sex Determination in Sciara, *Quart. Rev. Biol.*, vol. 6, pp. 306–312.

——— 1934. The Rôle of the Chromosome Sheath in Mitosis and Its Possible Relation to Phenomena of Mutation, *Proc. Nat. Acad. Sci.*, vol. 20, pp. 159–163.

——— 1935*a*. Structure of the Salivary Gland Chromosomes in Sciara, *Jour. Heredity*, vol. 26, pp. 177–188.

——— 1935*b*. Internal Structure of Salivary Gland Chromosomes in Sciara, *Jour. Heredity*, vol. 26, pp. 491–501.

——— 1937. Small Deficiencies and the Problem of Genetic Units in the Giant Chromosomes, *Genetics*, vol. 22, pp. 543–556.

——— 1938. Chromosome Behavior, Inheritance and Sex Determination in Sciara, *Amer. Naturalist*, vol. 72, no. 743.

——— and Elizabeth Gay Lawrence, 1936. Comparative Study of Salivary Gland Chromosomes in Diptera, *Amer. Naturalist*, vol. 70, p. 57.

——— ——— 1937. Studies on the Organization of the Giant Gland Chromosomes of Diptera, *Quart. Rev. Biol.*, vol. 12, pp. 135–151.

——— ——— 1938. Preliminary Observations on Sciara Hybrids, *Jour. Heredity*, vol. 29, pp. 179–186.

——— Mildred S. Moses, and Ella N. Hoppe, 1926. Chromosome Behavior and Genetic Behavior in Sciara (Diptera), I: Chromosome Behavior in the Spermatocyte Divisions, *Ztschr. f. ind. Abs. u. Vererb.*, vol. 42, pp. 237–270.

Muller, H. J., 1932. Further Studies on the Nature and Causes of Gene Mutations, *Proc. VI Internat. Cong. Genet.*, vol. 1, pp. 213–255.

——— 1935. The Origination of Chromatin Deficiencies as Minute Deletions Subject to Insertion Elsewhere, *Genetica*, vol. 17, pp. 237–252.

——— and M. S. Gershenson, 1935. Inert Regions of Chromosomes as the Temporary Products of Individual Genes, *Proc. Nat. Acad. Sci.*, vol. 21, pp. 69–75.

Painter, T. S., 1933. New Method for the Study of Chromosome Rearrangements and the Plotting of Chromosome Maps, *Science*, vol. 78, pp. 585–586.

Pätau, K., 1935. Chromosomenmorphologie bei Drosophila melanogaster und Drosophila simulans und ihre genetische Bedeutung, *Naturwiss.*, vol. 23, pp. 537–543.

Poulson, D. F., and C. W. Metz, 1938. Studies on the Nucleolus-forming Regions and Related Structures in the Giant Salivary Gland Chromosomes, *Jour. Morphol.*, vol. 63, pp. 363–395.

Smith-Stocking, Helen, 1936. Genetic Studies on Selective Segregation of Chromosomes in Sciara coprophila Lintner, *Genetics*, vol. 21, pp. 421–443.

Stadler, L. J., 1932. On the Genetic Nature of Induced Mutations in Plants, *Proc. VI Internat. Cong. Genet.*, vol. 1, pp. 274–294.

PLATE 1

All photomicrographs are from larval salivary gland cells.

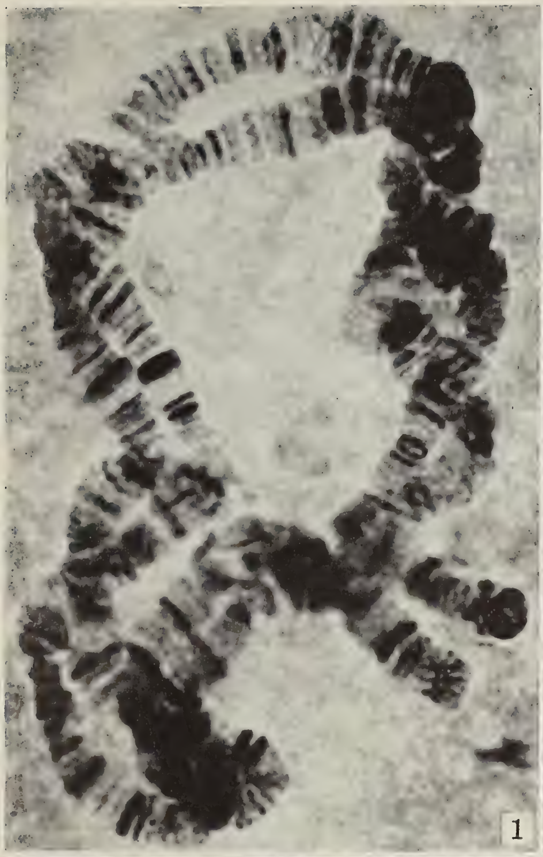
FIG. 1. The *X* chromosome pair in a hybrid between *Sciara ocellaris* and *S. reynoldsi*, showing the loosely paired association of the two corresponding or "homologous" members, with fusion of homologous parts in a few places. × 1700.

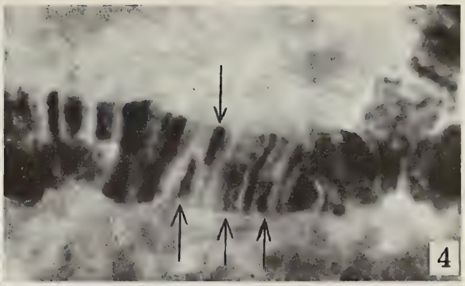
FIG. 2. Part of chromosome pair *A* in a hybrid between *S. ocellaris* and *S. reynoldsi*, showing general correspondence in pattern in the homologues near end 1, but a distinct, small difference at the disk marked 6. Another chromosome, extending in from the right, overlaps this pair. Compare with figure 3. *oc*, *ocellaris; re*, reynoldsi × 1700.

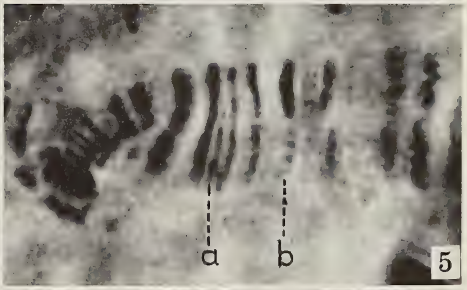
FIG. 3. End 1 of chromosome pair *A* in *S. ocellaris*, showing complete side-by-side fusion of the homologues except at the tip. A small difference in the structure of the two homologues is seen at 6. This appears to correspond accurately to the one in the species hybrid shown in figure 2. × 2500.

FIG. 4. Portion of the *X* chromosome pair in a hybrid between *S. ocellaris* and *S. reynoldsi*, showing three or four examples (indicated by arrows) of very small structural differences between the two members. These are distributed through a short region where in general the two chromosomes are alike. × 1700.

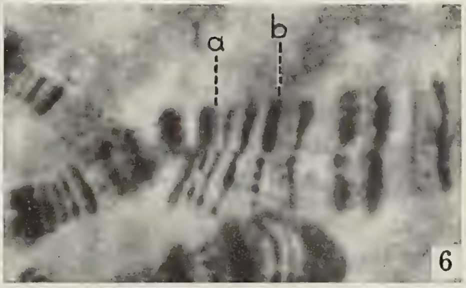
FIGS. 5, 6. End 1 of chromosome pair *A* of *S. ocellaris*, from different cells of one specimen, showing two very small structural differences between the two homologues (at *a* and *b*). These appear to be comparable to those found in the species hybrids. × 1700.


1

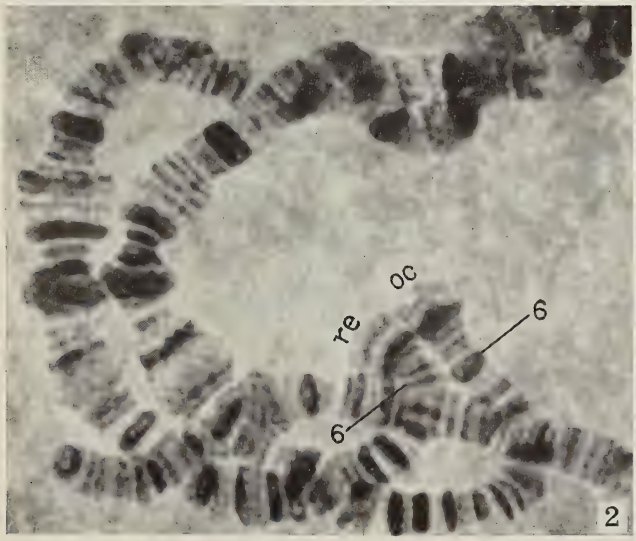

4



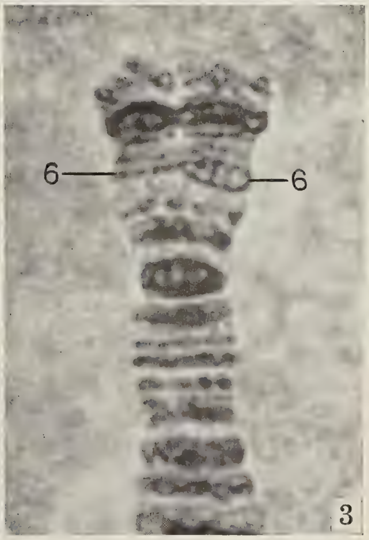
5



6



2



3

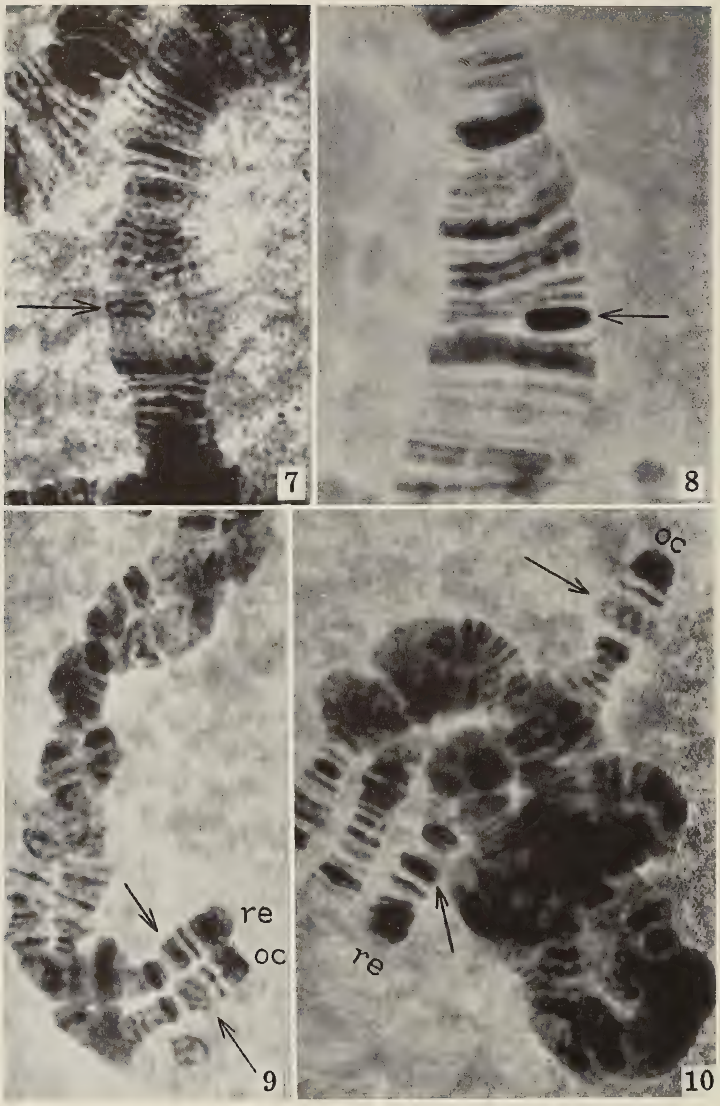


Fig. 7. A region toward the middle (from end 1) of chromosome pair *B* in *S. ocellaris*, showing a pair of heavy disks present in one homologue and apparently entirely absent from the other. × 1700.

Fig. 8. End 2 of chromosome pair *A*, showing a somewhat similar asymmetry. × 2500.

Figs. 9, 10. End 1 of chromosome pair *B* in hybrids between *S. ocellaris* and *S. reynoldsi*, showing what may represent a qualitative difference in apparently homologous disks or loci (indicated by arrows). At this location the disks in the *reynoldsi* chromosome appear as distinct cross-bands, while those in the *ocellaris* chromosome are disrupted by large achromatic droplets (see text). *oc*, *ocellaris; re*, *reynoldsi*. × 1700.

EIGHTEEN YEARS OF RESEARCH ON THE GENE

M. DEMEREC

Department of Genetics
The Carnegie Institution of Washington

THE problem which is now the focal point of all my research is not the original one on which I began my apprenticeship in genetics. It started as a casual interest, evolved into a hobby, and finally became the main objective of my work.

When I came to Cornell University as a graduate student, in 1919, I found studies on the genetics of maize at a very active stage. Professor R. A. Emerson was working intensively on linkage relationships, and at the same time carrying on numerous other studies covering many important phases of genetics. Professor C. B. Hutchinson was his right hand man, and he also was working on linkages and investigating a number of other problems. An integral part of the maize research group were we, graduate students including E. G. Anderson, W. H. Eyster, and L. A. Randolph.

At that time there was a great need for more workable characters which could be used for the genetic mapping of chromosomes. After many talks and discussions I was assigned the task of finding characters which were expressed in the early seedling stage. Emerson and Anderson placed at my disposal their collections of seed material, consisting of several thousands of ears of maize. Thus the work of growing an immense number of seedlings was begun, first testing self-pollinated ears from the collections and later testing the offspring from intercrosses between various newly found types. New types found in the inbred material were so numerous that the work had to be limited to chlorophyll characters. After four years of research about thirty genes affecting chlorophyll development were established.

During that period Anderson and I spent hours in the laboratory figuring out different solutions for problems arising from the experiments, in going over the data of accounts appearing in the periodicals, and in discussing the outstanding genetic hypotheses of that time. One of the fascinating topics of our discussions was the gene problem. It was a tantalizing thought that there were some sort of particles in the living cells, known to geneticists as genes, which were so powerful as to be responsible for the transmission of hereditary characteristics, and at the same time so minute and stable as to be beyond the reach of the cytological and experimental methods in use at that time. It was known then that genes have a high degree of individuality, that, on rare occasions, changes are induced in them by some unknown force, but that they are too small to be seen through the best of microscopes. During this period Morgan's hypothesis that genes are located in chromosomes in a linear order was being firmly established.

One of the outstanding questions in our minds was what we could do to find out more about genes, about their properties and the nature of their changes. This subject came up over and over again in our discussions. There was good reason for our intense interest in the gene problem since Emerson had a number of experiments under way in connection with his studies of pericarp variegation. He (Emerson, 1914) put forth the hypothesis interpreting that and other similar variegations frequently found in plants as being caused by unstable genes. Kernels of the variegated pericarp line show mosaic areas of white and red tissue. According to Emerson's interpretation, the gene for white pericarp color is in an unstable condition changing frequently into a gene responsible for red color. All red sectors appearing on ears of the variegated line of maize may be considered to be caused by such changes (mutations) occurring during the development of these ears. Thus we were at the very source of information about a frequently mutating gene, and we had at our disposal material showing a gene in a condition most favorable for experimental purposes.

I enjoyed working with maize but I began to feel that it would be advantageous to extend the scope of my work by studies on some other plants. Consequently I began to search for a plant which would be favorable for genetic research and would also satisfy both my main interest, which was the study of chlorophyll inheritance, and my hobby, which was the unstable gene problem. I approached various members of the Department of Horticulture who were raising seedlings of a large number of species of plants, and one day I was given a seedling of annual larkspur (*Delphinium ajacis*) showing white and green chlorophyll variegation. Since *Delphinium* is well adapted for genetic work, seeds of various garden varieties were collected and new studies begun. In the material so collected there were three variegated types affecting flower color and two affecting chlorophyll.

The work with *Delphinium* progressed and my interest in it increased appreciably since this material proved to be especially well adapted for the study of unstable genes. Four such genes were found and investigated. Although my main problem was still the inheritance of chlorophyll, my interest was shifting toward variegation studies. A strong stimulus in that direction came from a letter written in 1920 to Emerson by Anderson, who, by that time, had completed his studies at Cornell and held a position in the Department of Genetics of the Carnegie Institution of Washington at Cold Spring Harbor, New York. In order to explain frequent mutations in unstable genes he suggested that genes are not the ultimate units of heredity but are compound structures composed of still smaller units. In the case of variegated pericarp color he assumed that unstable genes contained two kinds of sub-gene particles, a white-determining kind and a red-determining kind, which by mere segregation formed either white- or red-determining genes.

This hypothesis, carrying the concept of the mechanistic structure of the gene to the extreme, was a definite forward step in gene research, since it opened a way for experimental approach to the problem. In the case of variegated pericarp, the red spots are of different sizes varying from very minute dots to large sectors covering a number of kernels. It seemed clear that the same mechanism was responsible for the origin of all these spots, and that the differences in size must be due to differences in the time of origin, namely, that large spots originated early in the development of the ear and small ones at a later stage. Thus by classifying spots according to their size and determining the frequency of each class it was possible to find out the approximate rate of change in the successive cell generations. Also by assuming that sub-gene particles assort freely it was possible to compute the theoretical expectancies for the rate of change at each division. Since that rate depends on the number of segregating particles within the larger unit, an estimate of the number of sub-genes within an unstable gene could be made.

There was a way open, therefore, to test experimentally these and other assumptions, and a new approach to the analysis of the gene became possible.

Anderson's hypothesis acted as a catalyzer for our variegation studies, and led to a reexamination of the available material. By that time Eyster had also left Cornell, but Anderson sent him a copy of his letter to Emerson, and they jointly undertook to determine the rate of mutations in the successive cell generations by using the material which they had available from other experiments (Anderson and Eyster, 1928). The first report describing this new approach to the problem of variegation was given at the Toronto (1921) meeting of the American Association for the Advancement of Science in a series of papers presented by Emerson, Anderson, Eyster, and myself (Emerson, Eyster, Anderson, and Demerec, 1922).

Upon completing my studies at Cornell in the spring of 1923 I became associated with C. W. Metz in the Department of Genetics of Carnegie Institution of Washington at Cold Spring Harbor. I worked with *Drosophila virilis* but brought my maize material with me in order to complete the chlorophyll studies, and the *Delphinium* material to continue the work on variegation. While trying to induce mutations by centrifuging, various chemicals, and certain other environmental factors, an unstable reddish body-color mutant was discovered in *Drosophila virilis* in January 1924 (Demerec, 1926, 1928). Unstable miniature wing character and unstable magenta eye color were found in November 1925 and June 1926 respectively (Demerec, 1926*a*, 1927). Several years later (February 1931) another unstable miniature was discovered (Demerec, 1933*a*). These four unstable genes are still the only genes in animals known to behave similarly to the variegated characters in plants, where more than fifty are known. No reason can be advanced as to why three of these extremely rare genes were discovered within a short period of time. A search for such genes was carried on for years afterward, not only in *D. virilis* but also in *D. melanogaster*, but proved unsuccessful.

If changes in unstable genes are caused by an assortment of smaller units within these larger gene units, then the frequencies observed experimentally should fall within the range of computed probabilities. Anderson succeeded in interesting Professor Sewall

Wright in this problem and he worked out theoretical expectancy curves for various conditions. The curves obtained by Anderson and Eyster on the mutation rate of the variegated pericarp gene showed a reasonable agreement with expectancy. However, there was one case which showed a striking instance of disagreement between observation and expectancy. Emerson (Emerson, Eyster, Anderson, and Demerec, 1922) observed that the mutation rate in the heterozygous material, carrying one unstable gene, was much higher than in the homozygous material, having two such genes.

With four unstable genes available in the *Delphinium* material it was possible to undertake a comparative study of similar problems on different genes, while the discovery of unstable genes in *Drosophila* opened up a way for a study of a number of problems for which plants could not be used to advantage. The results obtained with *Delphinium* (Demerec, 1931) indicated strongly that the mechanism responsible for variegation was not as simple as was assumed by the sub-gene or the genomere hypothesis, as it was called by Eyster (1924). Data on the mutability rate of unstable rose flower color showed that gene to be mutating with approximately the same rate throughout the twelve cell generations covered by these data. These observations did not agree with the theoretical expectancies based on Wright's calculations. According to Wright, in order to obtain the constant rate of change of 7^{-5} per cell generation which was observed in unstable rose, about 4000 genomeres would be required. With such a high number of components, thousands of cell generations would be needed before the constant rate of change would be approached. Therefore, a random assortment of genomeres could not give a horizontal line for the mutability curve under the biological conditions known to exist in *Delphinium.* Further evidence against the genomere hypothesis came from studies of the unstable gene for lavender flower color. It was found that the mutation rate of the lavender gene varies during ontogeny; namely, that it is high during the early embryonic development and again during the last few cell divisions occurring in flowers, but very low during the intermediate stages.

Studies with *Delphinium* raised serious doubts as to the validity of the genomere hypothesis. These doubts were further strengthened by the results obtained from the studies of unstable genes in *Drosophila virilis.* A restriction of instability to certain stages of ontogeny was found to be general in *Drosophila*. For example, the reddish-body gene is unstable only at the maturation division of heterozygous females (Demerec, 1928); the miniature gene is stable during early stages of embryonic development but becomes unstable later; one allel of that gene is unstable both in somatic and in germinal tissues, while another allel is unstable in somatic tissues only (Demerec, 1929*a*). It was also found that the rate of mutability may be greatly influenced by the genic composition of the individual and by the sex of flies. Five genes were isolated which affect the mutability rate of unstable miniature (Demerec, 1929, 1930, 1932). After several years of research with unstable genes of *Delphinium* and *Drosophila*, facts were accumulated which could be explained by the genomere hypothesis only if that hypothesis were supplemented by a number of assumptions. All these facts, however, were in good agreement with the alternative hypothesis visualizing unstable genes as chemically unstable units, and changes in unstable genes as chemical changes (Demerec, 1931, 1935*b*). All three unstable genes known at that time in *Drosophila virilis* originated in the laboratory as mutations from the wild-type, and all of them revert back to the wild-type, which is then stable. It is assumed that the original chemical change in the gene which gives rise to the unstable form is a reversible chemical reaction, the reverse change occurring under certain conditions of the cellular environment. Thus such a gene may be stable at one stage of ontogeny and unstable at another, and its rate of change may be affected by various other genes. With a few exceptions all unstable genes studied in plants behave in a similar manner. The majority of them are unstable in one direction only, the changes going from the recessive mutant form into the dominant wild-type form. The chemical change hypothesis, therefore, accounts without any difficulty for all facts observed experimentally.

In order to study the properties of a biological system which cannot be subjected to direct chemical analysis it is essential that this system be variable or still better that it be possible to induce that variability experimentally. The unstable gene material, characterized by high frequency of change, is more suitable for studies of the properties of the gene than the relatively stable gene material. Unfortunately, however, changes occurring in unstable genes are spontaneous changes and all attempts to modify them by external influences were unsuccessful. There appears to be no difference between stable and unstable genes in regard to the experimental induction of changes; both of them are very resistant to external factors (Demerec, 1932*a*).

Muller's (1927) discovery that X-rays are able to induce hereditary changes provided a welcome tool for studies of gene properties. When that discovery was made, the analysis of the behavior of unstable genes of *Delphinium* and *Drosophila* was well under way, and our plans were then reorganized to include studies of gene changes induced by X-ray treatment. One of our first X-ray experiments revealed that such treatment did not affect to any measureable degree the rate of change in the unstable genes of *Drosophila virilis*. Since it was known that X-rays are effective in increasing the rate of mutability of stable genes, this finding suggested that the reactions responsible for changes in unstable genes are not increased by X-rays in the same proportion as reactions responsible for changes in stable genes.

Using *Drosophila melanogaster* as material, a study was undertaken to determine the frequency and types of changes induced in 22 loci of the X-chromosome (Demerec, 1934*a*). Among 633,787 chances to detect a change, 156 changes were observed, indicating that the average frequency of change per locus was 0.0246 per cent. By dividing the frequency of changes in all the X-chromosome loci, which was found to be 12.5 per cent, by the average frequency of change per single locus, an estimate of the total number of loci present in the X-chromosome was obtained. This number is 500 for the X-chromosome and 2500 for all chromosomes, which agrees fairly well with estimates obtained by others by different lines of reasoning.

It is known that lethals, namely, gene changes which prevent the survival of the individuals which carry them, are the most frequent among the changes occurring spontaneously as well as among those induced by X-rays. A lethal change may occur in the loci where allels with a morphologically detectable effect are known as well as in the loci where such changes are unknown. In the experiment mentioned above, where changes were observed in certain selected loci, about 85 per cent of these changes were lethals, while only 15 per cent were viable. This and other similar observations raised several questions, such as, what are lethal changes, and also what is the difference between lethals and visible changes? Answers to these questions were suggested in the Carnegie Institution of Washington lecture on "What Is a Gene?" (Demerec, 1933), where a working hypothesis as to the nature of the gene was elaborated.

In this lecture a gene was defined as "a minute organic particle, capable of reproduction, located in a chromosome, and responsible for the transmission of a hereditary characteristic." There was ample experimental evidence to indicate that genes are submicroscopic and estimates made by various workers place the upper limit for the diameter of a gene between 10 and 70 mμ. Genes were visualized as single organic molecules and changes in genes as either changes or rearrangements within the molecule. Different allels of the same gene were looked upon as the detectable results of changes in different regions of a molecule and unstable genes were considered to be unstable chemical changes which revert back to the original form. It was suggested that the majority of lethals may be deficiencies, namely, losses of certain genes, and that a gene may be lost through mechanical elimination, physiological inactivation, or most likely by a failure to reproduce itself. Since the great majority of changes in any one locus are lethal, it would follow that the self-reproductive power of a gene is a very sensitive property which is readily affected. According to this concept a gene molecule can tolerate only a very slight change. Any radical change eliminates the gene from the gene complex and the elimination of certain genes has a lethal effect on the organism. This leads to the conclusion that a full complement of certain genes is required for the life of the organism and the evidence accumulating at that time indicated that even a single tissue cell may not be able to function properly without this full complement. Therefore, "If a whole complement of genes is necessary for an organism to live or even for a cell to function properly, the primary function of a gene is not the one by which we recognize it, viz. the determination of a visible characteristic, but it is the regulation of the life processes of the cell. The effects which different genes exert on different visible characteristics of an organism are only incidental to their primary function" (Demerec, 1933).

Experiments were devised to test the ideas expressed in this working hypothesis on the nature of the gene. The work on unstable miniatures of *Drosophila virilis* suggested that allels, namely, various forms of one gene, may be due to changes occurring in different molecular regions of that gene molecule. There are five allels known in the miniature locus differing from each other in their effect on the size of wings. Two of these, miniature-3 and miniature-5, are unstable. Moreover, each of these two unstable allels has its own series of forms, which are morphologically identical but differ in the degree of instability. In both allels "alpha" forms are unstable in both germinal and somatic tissues, "gamma" forms are unstable in somatic tissues only, and "beta" forms are stable in both types of tissues; also in both, miniature-3 and miniature-5, changes from one form into another occur with low frequency. Therefore, the allels of the miniature locus may be arranged into a two-dimensional series as follows:

$$\begin{tikzcd}
\text{mt-1} \arrow[r] \arrow[dr] & \text{mt-3 beta} \arrow[r] & \text{mt-2} \arrow[r] \arrow[dr] & \text{mt-5 beta} \arrow[r] & \text{mt-4} \\
& \text{mt-3 gamma} & & \text{mt-5 gamma} & \\
& \text{mt-3 alpha} & & \text{mt-5 alpha} &
\end{tikzcd}$$

Arrangement along the horizontal line is according to increasing size of wings and down the vertical line according to increasing degree of instability. It is evident that the relationship is closer between the forms in each of the two vertical groups than between the allels in the horizontal group. It could be expected that such a condition would arise if changes responsible for the horizontal series were located in different structural components of a molecule and changes responsible for the vertical series within the same component.

It should be possible to test this assumption experimentally, and an unsuccessful attempt was made to do it by studying changes in the white locus of *D. melanogaster*. In that locus more than a dozen distinct allels are known, four of which are red, eosin, apricot, and white. White can be obtained from any of the other allels. If it were true that a change in a certain region of the molecule gives the eosin gene, an independent change in another region gives apricot, and that a change in still another region gives white, then whites obtained from red, eosin, and apricot should be different from each other in that a reversion from white obtained from red should give red, white from eosin should give eosin, and so on.

Unfortunately, however, this experiment did not bring results since, although 47,000 offspring of X-rayed males were examined, reversions from white could not be obtained.

To investigate the importance of genes in the vital processes of cells another set of experiments was undertaken (Demerec, 1934, 1935). From the evidence available at that time it was inferred that a large proportion of lethals are deficiencies and that at least some of them are single gene deficiencies. That indicated that the presence of certain genes is essential in order that an organism may live. Two questions arose from this reasoning, namely, what happens to a small group of cells in which these genes essential for the life of the organism are missing, and what is the phenotype of the tissues deficient for a certain known gene? After several trials with various techniques, the technique discovered just at that time by Stern (1934) was adopted. Stern found that various Minute characters stimulate the segregation of X-chromosomes in somatic cells. Thus if each of the two X-chromosomes carries a gene through which it can be detected on the body of a fly, the presence of a Minute causes a frequent appearance of twin spots, one showing the character carried in one of the X-chromosomes and the other, the character carried in the other chromosome. Now, if one of the X-chromosomes in addition to a marker by which it can be recognized carries the deficiency which is being studied, and if both twin spots appear, this means that this deficiency is not lethal to the small group of cells forming the spot, but if only one of the twin spots shows up, this indicates that the deficiency is lethal. Such a condition is named cell-lethal, meaning a lethal which prevents the appearance of a small patch of tissue, in the present case in hypodermal cells of females for which tissue this method is applicable.

By this method more than 40 lethals, presumably deficiencies, affecting 15 different loci occupied by visible mutations of the X-chromosome were tested, and for all but two a cell-lethal condition was found. The two exceptions were a short deficiency affecting the yellow-achæte loci at the tip of the chromosome (Demerec and Hoover, 1936) and deficiencies affecting the cut locus. Since spots homozygous for the yellow deficiency are phenotypically yellow, this showed that the absence of the yellow gene has an effect similar to a change of the locus to an allel and gave an answer to the second question stated at the beginning of the work on this problem. Identical answers were obtained independently by Ephrussi (1934) and Muller (1935).

Experimental evidence that small deficiencies, presumably single gene deficiencies, may be cell-lethal strengthened the contention earlier expressed that genes play an important rôle in the vital processes of individual cells. This extended the application of Bridges' gene balance theory to single cells and suggested that even in higher organisms individual cells retain a certain degree of autonomy.

While experiments made with lethals affecting known loci indicated that the great majority of them are cell-lethal (13 out of 15), similar experiments conducted with random lethals showed only 10 out of 24 to be cell-lethals (Demerec, 1936). The higher frequency of cell-lethals among lethals affecting visible loci suggests that, on the average, these loci play a more significant rôle in the vital processes of the organism than do the others. According to their importance in vital processes loci can be grouped into several classes (Demerec, 1935*a*). These in the order of decreasing importance are as follows:

1. Heterozygous deficiency shows a dominant character, exaggerates mutant allel, and is cell-lethal when homozygous (Notch, Minutes).
2. Heterozygous deficiency has no dominant character but exaggerates mutant allel, homozygous deficiency is cell-lethal (forked, tan sable, tiny, rudimentary).
3. Heterozygous deficiency shows neither dominant effect nor exaggeration, homozygous is cell-lethal (dusky, miniature, fused).
4. Homozygous deficiency is lethal but not cell-lethal (cut).
5. Homozygous deficiency is viable and appears as a mutant character (yellow).
6. Homozygous deficiency does not show any effect as far as could be detected (Df 260–5).

With the discovery of the hereditary effects of X-rays many new problems opened up for the study of which *Drosophila* was unquestionably the most suitable material. The number of laboratories engaged in *Drosophila* research, large at that time, was becoming larger every year, so that the intimate contact which existed among the *Drosophila* workers of the past was being lost. Owing to this lack of contact much valuable material was not being fully utilized, and a great deal of effort was being duplicated. To remedy this situation a cooperative service for the exchange of information, called the Drosophila Information Service, was instituted in 1934 in collaboration with C. B. Bridges. From that time on mimeographed bulletins of 50 to 130 pages each have been issued twice a year. They contain information contributed by various *Drosophila* laboratories which is not ordinarily suitable for publication but which is of importance to workers active in *Drosophila* research. At present 79 *Drosophila* research laboratories with 306 workers are cooperating in this service by contributing material for the bulletins. Thus an exchange of information has been established through which collections of stocks available in different laboratories have become easily accessible to all, and which acts as a clearing house for discoveries of new mutants and linkages, for news on improvements in laboratory and cytological technique, and for advanced information about current experiments, and provides an up to date directory of *Drosophila* workers.

Soon after the discovery of the hereditary effects of X-rays, investigations conducted independently at several laboratories established that the effect obtained is proportional to the dosage used but that it is independent of the wave length. It was of considerable theoretical importance to have reliable data on wave-length tests especially within the range of soft rays. Since these rays are readily absorbed by living tissues, a refined technique had to be used in order to obtain reliable measurements of dosage, and only a physicist was competent to undertake such measurements. Therefore, in order to complete this and several other experiments involving X-rays it became necessary for geneticists to work in cooperation with physicists. An experiment with soft rays was conducted in cooperation with Fricke (Fricke and Demerec, 1937), and independently similar experiments were made by Timofeeff-Ressovsky and Zimmer (1935). These experiments showed that there was no difference in the frequency of hereditary changes induced by soft rays (2.2 Å) and by hard rays when similar dosages were applied. Fricke and Demerec concluded that these genetic changes are not produced by activations within a gene, but that they are probably in the nature of sensitized reactions induced by the transfer of energy from neighboring activated molecules.

It was shown by another set of experiments (Demerec, 1938) that biological factors may play an important rôle in regard to the frequency of hereditary changes induced by X-rays. Two strains of flies, Oregon-R and Swedish-b, were simultaneously treated with various dosages, and the Swedish-b flies gave a frequency of hereditary changes about twice as high as the Oregon-R flies.

Another field in gene research was opened up when Heitz and Bauer (1933) and Painter (1933) recognized that the salivary glands of *Drosophila* larvæ have very large chromosomes. They found, moreover, that these chromosomes show constant banding through which any section of their length can be recognized. Bridges, therefore (1935), prepared detailed maps showing individual salivary chromosome bands, and for the first time it became possible to make an exact cytological analysis of hereditary changes connected with gene deficiencies and chromosome rearrangements. Through studies of salivary chromosomes it has been possible to confirm the assumption that certain lethals are connected with deficiencies which can be detected cytologically. It has been found that a majority of lethals affecting visible loci (Demerec, not published) and a fair proportion of random lethals (Slizynski, 1938) are such cytological deficiencies. On the other hand, a deficiency has also been detected which does not affect either the viability or the phenotype of the fly (Demerec and Hoover, 1936). Moreover, today the position of a number of genes is determined to within a minute section of a chromosome since subsequent work (Mackensen, 1935; Demerec and Hoover, 1936; Slizynska, 1938; and others) has supported Painter's (1933) original suggestion that the bands detectable in these chromosomes may be correlated with gene loci.

X-ray radiation causes, in addition to gene changes, frequent chromosomal rearrangements. Since genes are located in chromosomes it was of interest to determine whether there was any connection between gene changes and chromosomal aberrations. Experiments showed (Demerec, 1937) an interesting relationship between these two phenomena. Of the 61 lethals affecting known loci of the X-chromosome, and induced by an X-ray treatment of 2500–3000 r-units, 41 per cent had chromosomal aberrations such as inversions and translocations where one break of the aberration coincided with the region in which the lethal change occurred. Furthermore, in this material, the frequency of aberrations increases with the increase of dosage. Of the 30 visible mutations induced in the same loci by a similar treatment none were connected with aberrations. It has been found also that none of the spontaneous lethals were associated with chromosomal aberrations. A cytological study of salivary chromosomes had indicated that a majority of lethals in visible loci are deficiencies. Now, the coincidence between a breakage point of a chromosomal aberration and the place where a deficiency has occurred indicates that these two processes may be induced by the same mechanism and that chemical changes producing small deficiencies may be responsible for breaks in chromosomes. When such a break occurs, the free ends thus produced have a tendency to join again. If only one break occurs in a nucleus, as probably happens in the case of spontaneous changes, broken ends can only reunite and there is no possibility for a rearrangement associated with a lethal to be formed. If, however, the frequency of breaks per nucleus is high, as in the case of high-dosage X-ray treatment, the opportunity for the origin of chromosomal rearrangements is also high and, therefore, lethals induced by such treatment will frequently coincide with such rearrangements. There is, however, another explanation to account for the coincidence between lethals and aberrations observed in this experiment. It might well be that when a chromosomal aberration occurs, the change in the position of genes near the breakage points affects the activity of these genes in such a way as to produce a lethal effect. In this case lethals associated with chromosomal aberrations may or may not be deficiencies. A detailed salivary chromosome study of the breakage regions of several lethal inversions (Hoover, 1937, 1938) indicates that at some breakage points deficiencies are evident, while at others detectable deficiencies do not occur.

Experimental evidence from a number of sources is accumulating which indicates a much closer relationship between adjacent genes and between genes and chromosomes than was formerly thought to exist. The evidence on the biological effects of gene deficiencies, which has been discussed already, showed that biologically a gene is a part of a balanced system and that the biological effect of a gene can be studied only in relation to all other genes. Similarly in studying the nature of genes a gene should be considered as a part of a chromosome. With this in view a comprehensive study of X-ray-induced chromosome breaks was undertaken (Bauer, Demerec, and Kaufmann, 1938). The salivary chromosomes of first-generation offspring from untreated females and treated males were studied for various chromosomal rearrangements (inversions, translocations, transpositions, and duplications) and the breakage points determined to within one division of Bridges' maps. Two regions can be distinguished on *Drosophila* chromosomes, one called euchromatic and the other heterochromatic. When compared with the total length, the heterochromatic regions in salivary gland cell chromosomes are much shorter than the heterochromatic regions in metaphase chromosomes of ganglia and other tissues. The study of 1038 breaks showed (1) that they are distributed at random throughout the euchromatic regions, with the possible exception of distal sections where breaks tend to be slightly more frequent; (2) that they are distributed at random among chromosomes of comparable lengths; (3) that two breaks are independent of each other; and (4) that they are more frequent in heterochromatic than in euchromatic regions of similar lengths. It was found that the frequency of breaks increased with the dosage, this increase significantly deviating from proportionality. The data thus obtained favor the assumption that chromosomal rearrangements are brought about through breaks and subsequent reunion rather than through fusions and subsequent breaks. The independence of two breaks of each other and their random distribution suggest that a different mechanism is responsible for reattachments connected with induced inversions and translocations than for those connected with crossing-over.

Observations on the random distribution of induced breaks again brought to the foreground the problem of the relationship between gene changes and chromosomal rearrangements by providing a new approach. These observations formed a solid foundation for reasoning that if a change in a gene is responsible for a break in the chromosome thread adjacent to it, and if chromosome breaks are distributed at random, then it should be expected that breaks would be more frequent in regions containing many genes than in regions containing only a few. It is known that heterochromatic regions, which are cytologically detectable both in salivary and in gonial chromosomes, generally contain very few genes in comparison with euchromatic regions. It is also known that the Y-chromosome is all heterochromatic. Therefore the distribution of breaks between the heterochromatic Y-chromosome and euchromatic chromosomes should show whether or not genes are responsible for breaks in chromosomes. The results of experiments showed that they are not responsible. The data indicated (Kaufmann and Demerec, 1937) that breaks occur with approximately equal frequency in the Y and in the autosomes.

The heterochromatic regions are much shorter in salivary chromosomes than in metaphase chromosomes; for instance, while Y in metaphase chromosomes is approximately of the same length as X, in salivary gland chromosomes X is about 50 times as long as Y. If the length of heterochromatic regions is determined from metaphase chromosomes, the number of breaks per unit length is similar in euchromatic and heterochromatic regions. This seems to indicate that the metaphase chromosomes approximate more closely than salivary gland chromosomes the proportional chromosome thread (chromonema) length of the two regions, and that the frequency of breaks per unit of length is the same in both. "If the frequency of breaks per unit length of chromonema is similar in two such different regions as the euchromatic and the heterochromatic, it would indicate that the structure of the chromonema is similar in both. This is equivalent to saying that the structure of the chromonemata is similar throughout the whole length of the chromosome, and this can be extended further to include all chromosomes. Chromosomes then can be visualized as composed of fibre-like chromonemata, which are structurally similar throughout the whole complex, and to which the genes are linked. Such a picture parallels closely the picture of a fibre protein molecule as outlined by Astbury (1934). The backbone of the protein molecule may correspond to the fundamental unit of a chromonema and a group of radicals attached to the backbone may correspond to what we call a gene" (Kaufmann and Demerec, 1937). Thus our studies on genes brought us into contact with protein biophysics and biochemistry.

In concluding this outline I shall try to present a balance sheet showing advances in our knowledge of the gene during the period considered here. At the beginning of that period genes were looked upon as independent units located in chromosomes in a linear order. They were considered by the majority of geneticists as having a specific effect on the determination of various characters, although a belief was expressed by some (Morgan, 1919) that the whole gene complex is instrumental in producing each and every part of the body. On the question of mutations the opinion was divided, some considering mutations as changes in genes, others as gene losses. Means for inducing changes in genes experimentally were unknown. At present genes are still thought of as structures located in a linear order in chromosomes. They are regarded as being present in every living cell of an organism. The view is now generally accepted that genes act as an integrated system which is responsible for the appearance of the organism as a whole, and, moreover, that the whole gene complex is responsible for the processes going on within individual cells. It is also known that the loss of certain genes is lethal to the organism and that such a loss may even prevent the appearance of a small patch of tissue. As a rule, mutations are looked upon as changes in genes, although evidence is available to indicate that some mutations may be due to gene losses and some may be produced by shifts in gene positions. During this period sufficient evidence has accumulated to make probable the assumption that changes in genes occur through chemical changes rather than through assortments of sub-gene particles. Various estimates placed the diameter of a gene at between 10 and 70 mμ, and the total number of genes for *Drosophila melanogaster* at between 2000 and 5000. One of the most far-reaching developments of this period was the discovery that X-rays and related radiations as well as heat treatments are able to induce changes in genes. Of almost equal importance for gene studies was the recognition that the salivary glands of Diptera have exceedingly large chromosomes showing a constant pattern of banding. With the help of that banding the exact position of a number of genes has been determined, the nature of changes investigated, and the similarities and differences in gene order between species studied (Dobzhansky and Tan, 1936). One of the recent important developments is represented by attempts to give a physicochemical interpretation to facts established by biological methods (Koltzoff, 1928; Demerec, 1933; Caspersson, 1936; Wrinch, 1936; Astbury and Bell, 1938) and to devise methods by which these interpretations can be experimentally tested.

When a large group of workers is engaged in research on one problem, as is the case in gene studies, the work of the whole group is usually so interlocked that it is not always possible to credit any one person for the solution of a certain problem. The work progresses toward its logical goal through the efforts of all. In this review of my work emphasis has been placed on those questions of the gene problem in which I am directly interested, and no attempt has been made to review or to evaluate the problem as a whole. Neither has any mention been made of a number of problems which developed in the course of this study and which were not related to the study of the gene.

LITERATURE CITED

Anderson, E. G., and W. H. Eyster, 1928. Pericarp Studies in Maize, III: The Frequency of Mutation in Variegated Maize Pericarp, *Genetics*, vol. 13, pp. 111–120.

Astbury, W. T., 1934. X-ray Studies of Protein Structure, *Cold Spring Harbor Symposia*, vol. 2, pp. 15–27.

——— and F. O. Bell, 1938. Some Recent Developments in the X-Ray Study of Proteins and Related Structures, *Cold Spring Harbor Symposia*, vol. 6. (In press.)

Bauer, H., M. Demerec, and B. P. Kaufmann, 1938. X-Ray Induced Chromosomal Alterations in Drosophila melanogaster, *Genetics*, vol. 23, pp. 610–630.

Bridges, C. B., 1935. Salivary Chromosome Maps, *Jour. Hered.*, vol. 26, pp. 60–64.

Caspersson, T., 1936. Über den chemischen Aufbau der Strukturen des Zellkernes, *Skandinavisches Arch. Physiol.*, vol. 73 (suppl.), pp. 1–154.

Demerec, M., 1926. Reddish—a Frequently "Mutating" Character in Drosophila virilis, *Proc. Nat. Acad. Sci.*, vol. 12, pp. 11–16.

——— 1926*a*. Miniature-alpha—a Second Frequently Mutating Character in Drosophila virilis, *Proc. Nat. Acad. Sci.*, vol. 12, pp. 687–690.

——— 1927. Magenta-alpha—a Third Frequently Mutating Character in Drosophila virilis, *Proc. Nat. Acad. Sci.*, vol. 13, pp. 249–253.

——— 1928. Mutable Characters of Drosophila virilis, I: Reddish-alpha Body Character, *Genetics*, vol. 13, pp. 359–388.

——— 1929. Genetic Factors Stimulating Mutability of the Miniature-gamma Wing Character of Drosophila virilis, *Proc. Nat. Acad. Sci.*, vol. 15, pp. 834–838.

——— 1929*a*. Changes in the Rate of Mutability of the Mutable Miniature Gene of Drosophila virilis, *Proc. Nat. Acad. Sci.*, vol. 15, pp. 870–876.

——— 1930. A Genetic Factor Affecting Germinal Mutability of Miniature-alpha Wing Character of Drosophila virilis. In The Laws of Life, pp. 45–56. Prague.

——— 1931. Behavior of Two Mutable Genes of Delphinium ajacis, *Jour. Gen.*, vol. 24, pp. 179–193.

Demerec, M., 1932. Rate of Instability of Miniature-3 gamma Gene of Drosophila virilis in the Males in the Homozygous and in the Heterozygous Females, *Proc. Nat. Acad. Sci.*, vol. 18, pp. 656–658.

——— 1932*a*. Effect of Temperature on the Rate of Change of the Unstable Miniature-3 gamma Gene of Drosophila virilis, *Proc. Nat. Acad. Sci.*, vol. 18, pp. 430–434.

——— 1933. What Is a Gene? *Jour. Hered.*, vol. 24, pp. 368–378.

——— 1933*a*. Miniature-5, a New Unstable Gene in Drosophila virilis, *Amer. Naturalist*, vol. 67, pp. 68–69.

——— 1934. Biological Action of Small Deficiencies of X-Chromosome of Drosophila melanogaster, *Proc. Nat. Acad. Sci.*, vol. 20, pp. 354–359.

——— 1934*a*. The Gene and Its Rôle in Ontogeny, *Cold Spring Harbor Symposia*, vol. 2, pp. 110–115.

——— 1935. Rôle of Genes in Evolution, *Amer. Naturalist*, vol. 69, pp. 125–138.

——— 1935*a*. Relative Importance of Various Genes to the Organism, *Science*, vol. 81, p. 420.

——— 1935*b*. Unstable Genes, *Bot. Rev.*, vol. 1, pp. 233–248.

——— 1936. Frequency of "Cell-lethals" among Lethals Obtained at Random in the X-Chromosome of Drosophila melanogaster, *Proc. Nat. Acad. Sci.*, vol. 22, pp. 350–354.

——— 1937. Relationship between Various Chromosomal Changes in Drosophila melanogaster, *Cytologia*, Jub. vol., pp. 1125–1132.

——— 1938. Hereditary Effects of X-Ray Radiation, *Radiology*, vol. 30, pp. 212–220.

——— and M. Hoover, 1936. Three Related X-Chromosome Deficiencies in Drosophila, *Jour. Hered.*, vol. 27, pp. 206–212.

Dobzhansky, Th., and C. C. Tan, 1936. Studies on Hybrid Sterility, III: A Comparison of the Gene Arrangement in Two Species, Drosophila pseudoobscura and Drosophila miranda, *Zeitschr. ind. Abst. Vererb.*, vol. 72, pp. 88–114.

Emerson, R. A., 1914. The Inheritance of a Recurring Somatic Variation in Variegated Ears of Maize, *Amer. Naturalist*, vol. 48, pp. 87–115.

——— W. H. Eyster, E. G. Anderson, and M. Demerec, 1922. Studies of Somatic Mutations in Variegated Maize Pericarp, *Anat. Rec.*, vol. 23, pp. 90–91.

Ephrussi, B., 1934. The Absence of Autonomy in the Development of the Effects of Certain Deficiencies in Drosophila melanogaster, *Proc. Nat. Acad. Sci.*, vol. 20, pp. 420–422.

Eyster, W. H., 1924. A Genetic Analysis of Variegation, *Genetics*, vol. 9, pp. 372–404.

Fricke, H., and M. Demerec, 1937. The Influence of Wave-length on Genetic Effects of X-Rays, *Proc. Nat. Acad. Sci.*, vol. 23, pp. 320–327.

Heitz, E., and H. Bauer, 1933. Beweise für die Chromosomennatur der Kernschleifen in den Knäuelkernen von Bibio hortulanus L., *Zeitschr. f. Zellforsch.*, vol. 17, pp. 67–82.

Hoover, M. E., 1937. A Tandem Inversion in Drosophila melanogaster, *Genetics*, vol. 22, pp. 634–640.

——— 1938. Cytogenetic Analysis of Nine Inversions in Drosophila melanogaster, *Zeitschr. ind. Abst. Vererb.*, vol. 74, pp. 420–434.

Kaufmann, B. P., and M. Demerec, 1937. Frequency of Induced Breaks in Chromosomes of Drosophila melanogaster, *Proc. Nat. Acad. Sci.*, vol. 23, pp. 484–488.

Koltzoff, N. K., 1928. Physikalisch-chemische Grundlage der Morphologie, *Biol. Zentralbl.*, vol. 48, pp. 345–369.

Mackensen, O., 1935. Locating Genes on Salivary Chromosomes, *Jour. Hered.*, vol. 26, pp. 163–174.

Morgan, T. H., 1919. The Physical Basis of Heredity, p. 305. Philadelphia, J. B. Lippincott Co.

Muller, H. J., 1927. Artificial Transmutation of the Gene, *Science*, vol. 66, pp. 84–87.

——— 1935. A Viable Two-Gene Deficiency, *Jour. Hered.*, vol. 26, pp. 469–478.

Painter, T. S., 1933. A New Method for the Study of Chromosome Rearrangements and the Plotting of Chromosome Maps, *Science*, vol. 78, pp. 585–586.

Slizynska, H., 1938. Salivary Chromosome Analysis of the White-Facet Region of Drosophila melanogaster, *Genetics*, vol. 23, pp. 291–299.

Slizynski, B. M., 1938. Salivary Chromosome Studies of Lethals in Drosophila melanogaster, *Genetics*, vol. 23, pp. 283–290.

Stern, C., 1934. Crossing-over and Segregation in Somatic Cells of Drosophila melanogaster, *Amer. Naturalist*, vol. 68, pp. 164–165.

Timofeeff-Ressovsky, N. W., and K. G. Zimmer, 1935. Wellenlangenunabhängigkeit der mutationsauslösenden Wirking der Röntgen- und Gammastrahlung bei Drosophila melanogaster, *Strahlentherapie*, vol. 54, pp. 265–278.

Wrinch, D. M., 1936. On the Molecular Structure of Chromosomes, *Protoplasma*, vol. 25, pp. 550–569.

FIFTEEN-YEAR BREEDING RECORDS OF $2n + 1$ TYPES IN DATURA STRAMONIUM

ALBERT F. BLAKESLEE
Director, Department of Genetics

AMOS G. AVERY
Department of Genetics
The Carnegie Institution of Washington

In 1915 in cultures of *Datura stramonium* grown at Storrs, Connecticut, there was discovered a type of peculiar form which was called the "Globe" mutant from the globose shape of its capsules. From breeding experiments carried on the next year at the Department of Genetics, the Globe type was found to be peculiar also in the way in which the Globe characteristics were transmitted in inheritance. When selfed or when used as a female in crosses to normal types, Globe plants appeared in about a quarter of the offspring. When the Globe type was used as a male parent, however, few if any Globe offspring resulted. Every year since the discovery of the Globe, cultures of *Datura stramonium* have been grown at the Department of Genetics. From time to time a new mutant type was identified in these cultures which has behaved in inheritance in a manner similar to the Globe.

In 1920, Belling (1; see bibliography at the end of the paper) furnished the chromosomal basis for these mutant types by showing that each contained a single extra chromosome. In that year also it was found possible for the first time to identify the particular chromosome in which a gene is located by means of trisomic ratios. In the same year tetrasomic inheritance from a tetraploid was first demonstrated. The year 1920 therefore forms the starting point of a concentrated study of chromosome behavior in relation to the morphology and physiology of *Datura.*

It was in 1920 that Dr. Merriam became President of the Carnegie Institution. It seems appropriate therefore in this volume in his honor to present records covering the major part of the years since 1920. The publications on *Datura* since 1920 which are listed in the bibliography are mostly joint papers. They show the extent to which the Datura studies have been a cooperative undertaking and bear witness to Dr. Merriam's continued encouragement and support during this fruitful 18-year period.

In 1920 we had in our collection a number of chromosomal forms each with an extra chromosome but had not yet recognized the different kinds of chromosomes which might be present as extras. Later (19), however, it was possible to distinguish primary $2n + 1$ (or simple trisomic) types (4), in which the extra chromosome is like the other two in the set of three, from secondary $2n + 1$ types in which the extra chromosome consists of the doubled half of one of the primaries. In addition there were identified tertiary trisomics (24) in which the extra consists of parts of two different primary chromosomes. The chromosomal distinction between these three types of trisomic mutants as well as their morphological and anatomical peculiarities and the method of their inheritance and transmission of genes have been discussed in a series of earlier papers listed in the bibliography and have been reported in the Annual Year Books of the Carnegie Institution.

When first discovered, the various primary, secondary, and to a less extent the tertiary $2n + 1$ types were distinguished by pet names, usually referring to some outstanding characteristic. About 8 years ago a terminology was adopted (39) whereby the two ends of each chromosome of our standard Line 1 are numbered like dominoes in consecutive order. The largest chromosome is called $1 \cdot 2$, the next in size $3 \cdot 4$, and so on to the smallest, which is $23 \cdot 24$. Our primary, which was once called Rolled (Rl), is now the $2n + 1 \cdot 2$ type, since it has the largest chromosome as the extra. Its two secondaries which were formerly called Polycarpic (Py) and Sugarloaf (Sg) respectively are now $2n + 1 \cdot 1$ and $2n + 2 \cdot 2$. The tertiary Maple (Mp) is now $2n + 4 \cdot 6$, since it is known to have the $\cdot 4$ and the $\cdot 6$ ends of the $3 \cdot 4$ and the $5 \cdot 6$ chromosomes respectively.

All the types discussed in the present paper belong to our highly inbred Line 1, except the tertiaries $2n + 1 \cdot 18$ and $2n + 2 \cdot 17$, and most of them belong to Line 1A which was inbred for 10 generations before 1922 when it was further purified by being passed through a haploid (10). All of the 12 primary $2n + 1$ types had been secured in the offspring of triploids before the beginning of the 15-year period under discussion. Of the secondaries, 10 were also known at that time but 3 of these were not in Line 1.

A word should be said regarding the routine method of handling the cultures from which the records to be discussed are taken. All the cultures are from selfs, with exception of 684 offspring from $2n + 1 \cdot 1$ and a large part of the tertiaries which were from female back-crosses to Line 1. Back-crossing has been found to give as many parental type offspring as selfing. Space prevents keeping all individuals of segregating cultures in the greenhouse until complete records can be taken and, moreover, some of the chromosomal types are difficult to identify under greenhouse conditions. Types have a better opportunity to express their distinguishing characteristics when grown to maturity in the field. This is especially true when more than a single extra chromosome is present. For these reasons the records of the present paper are confined to summer cultures which, except for a few individuals purposely kept in the greenhouse, have all been grown in the breeding field. Each spring in March and April selfed seeds from the $2n + 1$ types are counted and sown in seed pans in sterilized soil. Germination is usually relatively good and plants begin to appear from 7 to 10 days later dependent upon the temperature at which the pans are kept. Seedlings are potted up into 3-inch pots from 1 to 2 weeks later. If more seedlings show in the pan than are desired, care is always taken to pot up only those from one side of the pan so that the plants which are retained will form a random sample. If an inadequate number of seedlings have germinated by this time, the pan may be kept for two or three weeks longer and a second potting secured. There is a tendency for second pottings to give a somewhat higher proportion of $2n + 1$ types, as shown in table 1 in which the first and later pottings were recorded separately. The numbers are not large since the first and later pottings were not always kept separate in recording the adult types. The data, however, show consistently about twice as many chromosomal types in the later plantings both from primary and from secondary parents. This fact seems to indicate that the $2n + 1$ types are slower in germination than diploids. Miss Satina, moreover, has shown from the offspring of triploids (75) that one at least of the 12 chromosomes when it is present as an extra greatly delays germination of the seed. Germination delayed beyond the time when the seedlings were potted up may have lessened the proportion of $2n + 1$ types recorded. It would not have been practical, however, to have kept the seed pans till every last seed had germinated, since some seeds have been observed to germinate only after six months and more in seed pans. The error from discarding seed pans after only about a month's time would consist chiefly in decreasing the apparent viability of the plus 1 types and might be different for different chromosomes. Such an error would tend to act more or less uniformly throughout the 15 year period.

From 3 to 4 weeks after potting up the plants, they are ready to be transplanted into the field, which is not done before about May 15 when there is little further danger from frost. Plants which have grown too much before time for putting into the field are checked by cutting back the tips of their leaves. Seedlings which are too weak at this time to be safely put into the field are kept in the greenhouse for later determination. A certain proportion of the plants (about 10 per cent) are unrecordable from a number of vicissitudes such as early death from cutworms, infectious mosaic disease, etc. The records here are all based upon recordable plants. If each seedling had been given individual attention, there would probably have been a higher proportion of plus 1 types recorded, since they are less resistant to unfavorable conditions than are normals. In view of attempts at uniformity in treatment, however, and of the number of seasons represented, it is believed that the data are reasonably comparable and that the differences evident between the effects of different chromosomes are real.

TABLE 1. OFFSPRING OF PRIMARY (I) AND SECONDARY (II) $2n + 1$ TYPES IN FIRST AND LATER POTTINGS

Parents	First potting				Later pottings			
	Totals	Per cent parent type	Per cent related I	Per cent unrelated types	Totals	Per cent parent type	Per cent related I	Per cent unrelated types
Primaries	4499	19.76	—	.96	388	39.95	—	1.80
Secondaries	5033	13.39	4.41	.64	666	29.73	7.21	1.80

The cultures have been recorded usually by two persons working together—from 1924 to 1926 by the senior author and Gordon

Morrison, from 1926 to 1938 by one or both of the present authors. The $2n + 13 \cdot 14$ type for example, as well as some others, is not always readily distinguishable before fruiting. For this reason records have been checked several times during the growing season. Plants regarding which there was doubt have often had their chromosomal constitution determined cytologically, especially in later years of the 15-year period. Such plants have frequently been further tested by the character of their offspring.

The present paper reports on the breeding behavior of $2n + 1$ types for the last 15 years, beginning 1924. As will be pointed out in detail presently, the records show that the different extra chromosomes exert different influences upon their transmission to the next generation and upon the rate of mutation of unrelated $2n + 1$ types.

Such differences are in line with the effects of particular extra chromosomes upon the physiology and morphology of plus 1 types. Capsules have been found convenient to show morphological effects and a full series of capsules of all primary and secondary $2n + 1$ types and of the tertiaries $2n + 1 \cdot 18$, $2n + 2 \cdot 17$, $2n + 1 \cdot 9$ and $2n + 2 \cdot 5$ have been given in a recent publication (53). A capsule of the tertiary $2n + 4 \cdot 6$ is pictured in an earlier publication (29). Photographs of a complete series of seedlings of the simple trisomic forms have not heretofore been published. Those of the forms discussed in the present paper are shown in plates 1 to 5. They were photographed in 3-inch pots at or slightly before the stage at which they are transplanted to the field. Under each is given the original laboratory nickname and a numbered model of the extra chromosome responsible for the peculiarities of the type. It is evident that different extra chromosomes bring about recognizable differences in the appearance of the seedlings of primaries, secondaries and tertiaries.

The $2n + 1 \cdot 1$ seedling has very narrow leaves due to factors in the $\cdot 1$ half, those of its related secondary $2n + 2 \cdot 2$ are relatively broad, due to factors in the $\cdot 2$ half, while those of their primary $2n + 1 \cdot 2$ are intermediate in width due to factors in both halves. The general rule (to which there are exceptions) that characters of primaries are intermediate between those of their two secondaries is also shown by the $2n + 5 \cdot 6$, $2n + 9 \cdot 10$ and less well by the $2n + 13 \cdot 14$ groups (plates 1 and 2). The influence of extra parts of different chromosomes is shown in the tertiaries. Thus the influence of the ·1 half, which tends to be dominant in combinations, is seen in the $2n + 1 \cdot 9$ and the $2n + 1 \cdot 18$ types. The leaves of both are somewhat broadened by the other component of the tertiary chromosome. In the $2n + 2 \cdot 17$ the leaves are characteristically depressed at the base like those in $2n + 17 \cdot 17$ but their width is increased due to the ·2 half.

In view of the different effects brought about by different chromosomes upon morphological characters such as those seen in the seedlings, it is not surprising that they should be responsible also for differences in viability and mutation rate which the present study discloses.

Table 2 presents the raw data for the 15-year period. All the offspring recorded are from selfed parents or in relatively few cases included in the table are from female back-crosses to Line 1 diploids. The number of seed parents listed in the second column range from 14 in the case of the secondary $2n + 1 \cdot 1$, which is of weak growth and difficult to keep alive for the production of seed, to 43 in the case of the primary $2n + 19 \cdot 20$, which is of adequate vigor when full grown but which requires extensive plantings to ensure its appearance. The total recordable plants for each type, except in four cases, number two thousand or more individuals. The total number of recordable plants shown in the table is over ninety-one thousand.

The primary and secondary $2n + 1$ types are listed in the first column as parents and in the body of the table their offspring are shown under the appropriate headings. The secondaries are represented on either side of their own primaries. In cases in which one or both of the secondaries are missing, the vacant place is represented by a dash. Offspring like the parental type are printed in bold face type. Thus of 2049 recordable offspring from the primary $2n + 1 \cdot 2$, there were 1780 normal diploids, 213 $2n + 1 \cdot 2$ types like the parent, 1 individual of its secondary $2n + 1 \cdot 1$, and 5 individuals of its other secondary $2n + 2 \cdot 2$. In addition the $2n + 1 \cdot 2$ parent threw the following unrelated chromosomal types: 1 plant of $2n + 3 \cdot 4$, 3 plants of $2n + 5 \cdot 6$, 3 of $2n + 9 \cdot 10$, etc., as well as 23 tetraploids. The secondary $2n + 2 \cdot 2$ threw, out of 1857 offspring, 1424 normals, 332 parental types and 75 plants of the related primary $2n + 1 \cdot 2$. As will be noted by following the diagonal line of figures including the parental type offspring in bold face type, the secondaries throw a considerable percentage of their related primaries except $2n + 19 \cdot 19$, while primaries throw their own secondaries very rarely though more frequently than do unrelated primaries.

In table 2 the balanced types $1n$, $3n$, and $4n$ are represented, but for a number of reasons they, as well as modifications of these types, are not used in the tabulations in later tables which have to do with diploids and modified diploids. Spontaneous tetraploids when they first appear are generally mixochimeras (72), with leaves so malformed and roughened that the presence of extra chromosomes or deficiencies would be difficult to recognize. It would also be unfair to include haploids so far as the spontaneous appearance of types with extra chromosomes is concerned, since, though both primary and secondary $1n + 1$ types have been identified (67), they are extremely weak in growth and would be selected against in comparison with unmodified haploids.

The next to the last column includes the double $2n + 1 + 1$ types in which one of the two extra chromosomes is like that in the parental type. Thus the secondary $2n + 5 \cdot 5$ threw the double trisomics $2n + 5 \cdot 5 + 13 \cdot 14$, $2n + 5 \cdot 5 + 19 \cdot 20$ and $2n + 5 \cdot 5 + 21 \cdot 22$. Such double types are not counted in the body of this table but in later tabulations are credited to the types represented by each extra chromosome. Thus these three types with $5 \cdot 5$ extra chromosomes are added to the 702 $2n + 5 \cdot 5$ offspring making a total of 705 offspring like the parental type; and the same plants which also have the $13 \cdot 14$, $19 \cdot 20$, and $21 \cdot 22$ chromosomes are added to the simple trisomic offspring which have these chromosomes as extras. The $2n + 13 \cdot 14$ primary has thrown three simple tetrasomic $2n + (13 \cdot 14)_2$ types. Similarly the tetrasomics $2n + (15 \cdot 16)_2$ and $2n + (21 \cdot 22)_2$ have been thrown respectively 7 and 12 times by their simple trisomic parents. These 3 primary chromosomes may be transmitted to a very limited extent through the pollen. It is assumed that such tetrasomic ($2n + 2$) types have come from $1n + 1$ eggs fertilized by similar $1n + 1$ sperms. In consequence each is counted only once in later tabulations in which it has been assumed that no pollen transmission of extra chromosomes has taken place. These assumptions are not necessarily justified. The table shows a $2n + (21 \cdot 22)_2$ tetrasomic from an

TABLE 2. OFFSPRING OF PRIMARY (I), SEC

Summer records from 1924 to 1938. All plants recorded from 2*n*, I, and II parents, except 1421 offspri II's are from selfings except 331 offspring of 2*n* + 1·1 grown in 1924 and 353 in 1930 which were from fema 2*n* + 2·5 were in Line 1. The extra chromosomes in 2*n* + 1·18 and 2*n* + 2·17 were obtained from Prime T in the table were taken. Offspring from the III's were obtained from both selfs and female back-crosses. D in later tables. For such tabulations 1*n*, 3*n*, and 4*n* balanced types and modifications of such types are excl

Parent type	Number of parents	Plants potted	Dead and indeterminate	Recordable	2*n*	Types of 2*n* + 1 offspring designated by extra c																					
						1·1	1·2	2·2	3·3	3·4	—	5·5	5·6	6·6	7·7	7·8	—	9·9	9·10	10·10	11·11	11·12	—	13·13	13·14	14·14	15·15
2*n*	544	21780	1525	20255	20172		6						3						4						6		
2*n* +																											
1·1	14	2021	65	1956	1857	48	47	—					1												2		
1·2	17	2367	318	2049	1780	1	213	5		1			3						3			1			7		
2·2	17	2141	284	1857	1424	—	75	332	1				1						2						1		
3·3	28	2740	331	2409	2145		2		146	87			2						1			1			4	1	
3·4	18	2368	279	2089	1634				—	452																	
—																											
5·5	20	2594	338	2256	1400		5			2		702	120	—					2						2		
5·6	19	2662	295	2367	1591		2					2	725	—					5					1	6		1
6·6	15	2783	445	2338	1762			1				—	140	403													
7·7	23	3066	304	2762	1931		1								764	51			1								
7·8	17	2332	252	2080	1865					2					—	208						1			2		
—																											
9·9	21	2503	266	2237	1659		2					1	2			1		359	187	—		1			7		
9·10	18	2461	301	2160	1454		1						3					1	686	—		1			2		
10·10	20	2403	405	1998	1361		4						3					—	48	559		2			6		
11·11	20	2518	247	2271	1827		3						2						1		397	27					
11·12	18	2549	321	2228	1716		3						1	1					5		2	491			4		
—																											
13·13	15	2460	425	2035	1704		2						2						1					205	114	—	
13·14	17	2360	327	2033	1451		2						5			2		1	3			1		—	538	—	
14·14	16	2682	310	2372	1812		3						4						3					—	71	468	
15·15	20	2517	268	2249	1668		2			1			1						2						5		457
15·16	17	2598	320	2278	1788		1			1			4			1			4						3		1
—																											
17·17	16	2520	276	2244	1808		3						1												3		
17·18	20	2380	240	2140	1565		1			1			3									1			3		
—																											
19·19	22	3229	365	2864	1996		4			2			4						1						1		
19·20	43	5438	680	4758	4498	1	11		1	2		1	18						8		1	2			28		
—																											
—																											
21·22	18	2600	260	2340	1626					2									1								
—																											
—																											
23·24	17	2292	248	2044	1371		1																				
—																											
1·9	17	3028	539	2489	2072	1	14									2		2	25			1					
2·5	18	2447	358	2089	1436		16			1		1	18						1						5		
4·6	20	2720	248	2472	1714					1		1	3														
1·18	17	1940	146	1794	1486		4												1								
2·17	22	2715	252	2463	1939		11	1											1								

RY (II) AND TERTIARY (III) $2n + 1$ TYPES

$2n + 1\cdot 1$ were in L1A which originated from a haploid within Line 1. All the offspring of $2n$'s, I's, and
-crosses to L1A. Of the III types the offspring of $2n + 4\cdot 6$ were in Line 1A, those of $2n + 1\cdot 9$ and
ut these types had been purified by repeated female back-crosses to Line 1 or Line 1A before the records
$+ 1 + 1$ types are not included under $2n + 1$ types in this table but are counted twice for tabulations
ffspring like the parental type are printed in bold face type.

sne																	
7·17	17·18	—	19·19	19·20	—	—	21·22	—	—	23·24	—	$1n$	$3n$	$4n$	Parent type +		Other types
	4		1	1			3					3		48			
										1							
	5						4			3				23			
							1					1		18			
	1						3			2				11			$2n + 4\cdot 6$.
							1					1		1			
1	3						2							12	13·14; 19·20; 21·22.		$2n + 5\cdot 6 + 13\cdot 14$.
				1			2			4				18	2·2; 11·12; (3) 13·14.		
1							1					1		29			
							2					1		9	21·22; 23·24.		
							1							1			
	2						7			3		1		2			$2n + 9\cdot 10 + 5\cdot 6$; $2n + 9\cdot 10 + 21\cdot 22$; $2n + 5\cdot 6 + 13\cdot 14$.
	1						3			2				3	1·2; 5·6; 21·22.		
	2						1			1				1	5·6; 13·14; 17·18; 21·22.		$2n + 9\cdot 10 + 21\cdot 22$.
										2				8	9·10; 21·22.		$2n + \cdot 11$.
														1	13·14; 17·18		$2n - 1\cdot 2$.
										1				2			$2n + 9\cdot 10 + 13\cdot 14$; $1n + 13\cdot 14$; $3n + 13\cdot 13$.
	3			1			7			2		1		7	(3) 13·14; 14·14; 23·24.		$2n + 11\cdot 12 + 21\cdot 22$.
	2						4					1			13·14; 15·16; 23·24.		
							1			1				32	3·4; (2) 15·16.		
	2						5							2	(7) 15·16; 23·24.		
275	146											1		7			
1	558						2					2		2	15·16		
	2		850	2						1							$2n + 5\cdot 6 + 13\cdot 14$.
	3		7	141			18			4		1	1	3	15·16		$2n + 5\cdot 6 + 9\cdot 10$; $2n + (21\cdot 22)_2$.
	1						686							11	(12) 21·22		$2n + 13\cdot 14 + 15\cdot 16$.
										665				6	21·22.		
															Parent type	Parent type +	
							1					2			369		
							1			1					601	9·10; (2) 21·22	$2n + 5\cdot 6 + 9\cdot 10$; $2n + 5\cdot 6 + 13\cdot 14$.
												1		13	739		
	9						1							4	288		
	2											1		2	506		

TABLE 2. OFFSPRING OF PRIMARY (I), SECONDARY (II) AND TERTIARY (III) $2n + 1$ TYPES

Summer records from 1924 to 1938. All plants recorded from $2n$, I, and II parents, except 1421 offspring from $2n + 1 \cdot 1$ were in L1A which originated from a haploid within Line 1. All the offspring of $2n$'s, I's, and II's are from selfings except 331 offspring of $2n + 1 \cdot 1$ grown in 1924 and 353 in 1930 which were from female back-crosses to L1A. Of the III types the offspring of $2n + 4 \cdot 6$ were in Line 1A, those of $2n + 1 \cdot 9$ and $2n + 2 \cdot 5$ were in Line 1. The extra chromosomes in $2n + 1 \cdot 18$ and $2n + 2 \cdot 17$ were obtained from Prime Type 2, but these types had been purified by repeated female back-crosses to Line 1 or Line 1A before the records in the table were taken. Offspring from the III's were obtained from both selfs and female back-crosses. Double $2n + 1 + 1$ types are not included under $2n + 1$ types in this table but are counted twice for tabulations in later tables. For such tabulations $1n$, $3n$, and $4n$ balanced types and modifications of such types are excluded. Offspring like the parental type are printed in bold face type.

Parent type	Number of parents	Plants potted	Dead and in-determinate	Recordable	$2n$	Types of $2n + 1$ offspring designated by extra chromosome																																				$1n$	$3n$	$4n$	Parent type +	Other types
						1·1	1·2	2·2	3·3	3·4	—	5·5	5·6	6·6	7·7	7·8	—	9·9	9·10	10·10	11·11	11·12	—	13·13	13·14	14·14	15·15	15·16	—	17·17	17·18	—	19·19	19·20	—	—	21·22	—	—	23·24	—					
$2n$	544	21780	1525	20255	20172		6						3						4						6			4			4		1	1			3					3		48		
$2n$ +																																														
1·1	14	2021	65	1956	1857	**48**	47	—					1												2															1						
1·2	17	2367	318	2049	1780	1	**213**	5		1			3						3			1			7						5						4			3				23		
2·2	17	2141	284	1857	1424	—	75	**332**	1				1						2						1			1									1					1		18		
3·3	28	2740	331	2409	2145		2		**146**	87			2						1			1			4	1		2			1						3			2				11		$2n$ + 4·6.
3·4	18	2368	279	2089	1634				—	**452**																											1					1		1		
—																																														
5·5	20	2594	338	2256	1400		5			2		**702**	120	—					2						2			1		1	3						2							12	13·14; 19·20; 21·22.	$2n$ + 5·6 + 13·14.
5·6	19	2662	295	2367	1591		2					2	**725**	—					5					1	6		1	4						1			2			4				18	2·2; 11·12; (3) 13·14.	
6·6	15	2783	445	2338	1762			1				—	140	**403**																1							1					1		29		
7·7	23	3066	304	2762	1931		1								**764**	51			1																		2					1		9	21·22; 23·24.	
7·8	17	2332	252	2080	1865					2					—	**208**						1			2												1							1		
—																																														
9·9	21	2503	266	2237	1659		2					1	2			1		**359**	187	—		1			7						2						7			3		1		2		$2n$ + 9·10 + 5·6; $2n$ + 9·10 + 21·22; $2n$ + 5·6 + 13·14.
9·10	18	2461	301	2160	1454		1						3					1	**686**	—		1			2						1						3			2				3	1·2; 5·6; 21·22.	
10·10	20	2403	405	1998	1361		4						3					—	48	**559**		2			6			5			2						1			1				1	5·6; 13·14; 17·18; 21·22.	$2n$ + 9·10 + 21·22.
11·11	20	2518	247	2271	1827		3						2						1		**397**	27						1												2				8	9·10; 21·22.	$2n$ + ·11.
11·12	18	2549	321	2228	1716		3						1	1					5		2	**491**			4			1																1	13·14; 17·18	$2n$ − 1·2.
—																																														
13·13	15	2460	425	2035	1704		2						2						1					**205**	114	—		1												1				2		$2n$ + 9·10 + 13·14; $1n$ + 13·14; $3n$ + 13·13.
13·14	17	2360	327	2033	1451		2						5			2		1	3			1		—	**538**	—		3			3			1			7			2		1		7	(3) 13·14; 14·14; 23·24.	$2n$ + 11·12 + 21·22.
14·14	16	2682	310	2372	1812		3						4						3					—	71	**468**		1			2						4					1			13·14; 15·16; 23·24.	
15·15	20	2517	268	2249	1668		2			1			1						2						5		**457**	76									1			1				32	3·4; (2) 15·16.	
15·16	17	2598	320	2278	1788		1			1			4			1			4						3		1	**458**			2						5							2	(7) 15·16; 23·24.	
—																																														
17·17	16	2520	276	2244	1808		3						1												3					**275**	146											1		7		
17·18	20	2380	240	2140	1565		1			1			3									1			3					1	**558**						2					2		2	15·16	
—																																														
19·19	22	3229	365	2864	1996		4			2			4						1						1						2		**850**	2						1						$2n$ + 5·6 + 13·14.
19·20	43	5438	680	4758	4498	1	11		1	2		1	18						8		1	2			28			6			3		7	**141**			18			4		1	1	3	15·16	$2n$ + 5·6 + 9·10; $2n$ + $(21 \cdot 22)_2$.
—																																														
—																																														
21·22	18	2600	260	2340	1626					2									1												1						**686**							11	(12) 21·22	$2n$ + 13·14 + 15·16.
—																																														
—																																														
23·24	17	2292	248	2044	1371		1																																	**665**				6	21·22.	
—																																														
																																													Parent type | Parent type +	
1·9	17	3028	539	2489	2072	1	14									2		2	25			1															1					2			369	
2·5	18	2447	358	2089	1436		16			1		1	18						1						5			3									1			1					601 | 9·10; (2) 21·22	$2n$ + 5·6 + 9·10; $2n$ + 5·6 + 13·14.
4·6	20	2720	248	2472	1714					1		1	3																													1		13	739	
1·18	17	1940	146	1794	1486		4												1									1			9						1							4	288	
2·17	22	2715	252	2463	1939		11	1											1												2											1		2	506	

unrelated $2n + 19 \cdot 20$ parent and the same tetrasomic has appeared in offspring from selfing other unrelated trisomics as well as from selfing $2n$ parents in cultures not recorded in this table. Adequate evidence is not available regarding the behavior of chromosomes in all the divisions of the female gametophyte. If non-disjunction were known to occur in the second meiotic division following non-disjunction in the first, an alternative method for the production of tetrasomics would be available. In a number of cases selfs of trisomics have been found to throw fewer rather than more parental types in the offspring than female back-crosses. This was found to be true when the extra chromosome is transmitted through the pollen (3) as well as when it cannot be pollen transmitted (17). It is safe to say that the method of analysis of female gametes by selfing appears not to be a source of error in giving more parental types of offspring than back-crosses because of pollen transmission of extra chromosomes.

The last column labelled "other types" shows double trisomics unrelated to the parental type. Each of these is counted in later tables as two simple trisomic offspring. In the same column are listed a fragment type $2n + \cdot 11$ and a deficiency, $2n - 1 \cdot 2$. In later tabulations in which non-disjunction and segregation are involved these two types have been omitted. The deficiency probably occurred at an early stage in development of the seedling since evidence is good that gametophytes of Datura are not viable if any of the 12 chromosomes are missing. The $1 \cdot 2$ chromosome is the one most frequently eliminated to form $2n - 1$ sectors, of which we usually have observed several in our cultures each year. The $2n - 1$ plant listed in the table was selfed and gave 104 recordable offspring, of which 103 were $2n$ and 1 was a $2n + 5 \cdot 6$. Offspring from $2n - 1$ sectors in all cases failed to transmit the deficiency.

The data which have been presented in table 2 may be analyzed in several ways. They will be used primarily to throw light on three effects of specific extra chromosomes:

1. Inheritance of the extra chromosome through segregation.
2. The $2n + 1$ types as mutation producing machines of greater or less efficiency in bringing about the appearance of unrelated chromosomal types.
3. The origin of $2n + 1$ types as products of chromosomal mutation.

INHERITANCE OF EXTRA CHROMOSOMES

As has just been explained, the inheritance through the pollen is not considered in the present tabulations. Our breeding records give evidence for pollen transmission of extra chromosomes only in the case of 11·12, 13·14, 15·16, and 21·22 primary chromosomes and the 2·2 secondary chromosome of the types listed in table 2. Buchholz (46) has shown that pollen tubes with other extra chromosomes fail to function for a number of reasons. Pollen tubes containing as extras any of the five chromosomes just mentioned are considerably slower in growth than normals. Even though the transmission of these extra chromosomes may be greatly increased by special environmental conditions (45) it is clear that under ordinary conditions these extra chromosomes are inherited to only a very limited extent. It has been shown that in reduction divisions of a $2n + 1$ plant the chromosomes of pollen mother cells segregate in such a way that equal numbers of n and of $n + 1$ daughter cells result. Presumably a similar condition holds for megaspore mother cells. One would expect therefore 50 per cent $2n$ and 50 per cent $2n + 1$ offspring from a primary trisomic parent. That so high a proportion of parental types is never realized in the offspring must be due to elimination of the extra chromosome or to elimination of cells containing this chromosome during some stage or stages of development of the gametophyte or zygote. Such an elimination has been shown for extra chromosomes in the female gametophyte (71) and in the zygote (75) of triploids. Table 3 shows in percentages the different types of offspring thrown by primaries and secondaries. The unrelated types will be discussed in the section on mutations. From table 3 it is evident that there are great differences in the extent to which the different extra chromosomes are transmitted. The smallest primary chromosome 23·24 takes first rank in viability, being transmitted to 32.7 per cent of its offspring, while the third smallest chromosome 19·20 ranks 12th in viability with transmission to only 3 per cent of its offspring. Of the 14 known secondaries, the 5·5 ranks first in viability, with a transmission of 31.4 per cent, and the 1·1 chromosome ranks 14th with a transmission of only 2.5 per cent.

The relative viabilities of primaries and secondaries are also

TABLE 3. OFFSPRING OF PRIMARY (I) AND SECONDARY (II) $2n + 1$ TYPES

Parents $2n$ +	Viability rank		Per cent $2n + 1$ offspring					Per cent $2n$ offspring	Totals
			Parent type		Related types		Unrelated types		
	I	II	I	II	I	II			
1·1		14		2.45	2.40		0.20	94.94	1956
1·2	10		10.51			0.29	1.33	87.86	2026
2·2		7		18.06	4.08		0.38	77.48	1838
3·3		13		6.09	3.63		0.79	89.45	2398
3·4	8		21.66			0.0	0.05	78.29	2087
5·5		1		31·42	5·39		0.98	62.39	2244
5·6	3		31.08			0.09	1.32	67.73	2349
6·6		9		17.46	6.07		0.13	76.34	2308
7·7		4		27.83	1.85		0.22	70.17	2752
7·8	11		10.00			0.0	0.29	89.71	2079
9·9		10		16.07	8.46		1.34	74.26	2234
9·10	2		31.94			0.05	0.74	67.41	2157
10·10		3		28.19	2.45		1.45	68.15	1997
11·11		8		17.63	1.19		0.49	80.73	2263
11·12	7		22.14			0.09	0.76	77.05	2227
13·13		12		10.09	5.66		0.39	83.90	2031
13·14	5		26.81			0.05	1.63	71.65	2025
14·14		6		19.87	3.04		0.80	76.42	2371
15·15		5		20.75	3.52		0.63	75.24	2217
15·16	9		20.47			0.04	0.97	78.56	2276
17·17		11		12.30	6.53		0.31	80.86	2236
17·18	6		26.17			0.05	0.56	73.27	2136
19·19		2		29.68	0.07		0.59	69.69	2864
19·20	12		2.99			0.15	2.27	94.63	4753
21·22	4		29.97				0.26	69.82	2329
23·24	1		32·68				0.10	67.27	2038
Averages.................			22.20	18.42	3.88	0.08	0.73	77.05	60191
$2n$..							0.16	99.84	20204

shown by the graphs of figure 1. It is evident that viability of these trisomics is not conditioned by size of the chromosome which is extra. In some cases the primary may have better viability than its secondary (e.g., the 3·4 chromosome) while in other cases the primary is less viable (e.g., the 19·20 chromosome). Of the four primaries with both secondaries known, the 1·2 is intermediate in viability, the 5·6 has the same viability as one of its secondaries and two (the 9·10 and the 13·14) have greater viability than either secondary. If all secondaries were known, it might be found that there is a tendency for the primary either to be intermediate in viability between its two secondaries or for the primary to resemble in viability only one of its secondaries.

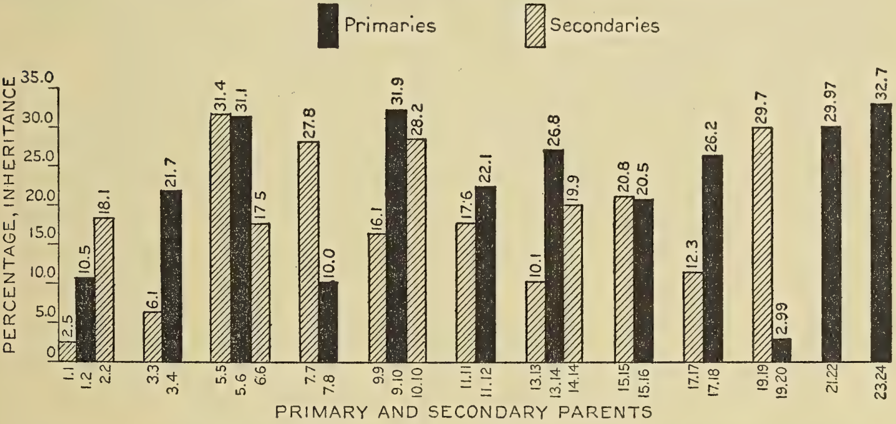


FIG. 1. Viability of primary and secondary $2n + 1$ types. The $2n + 1$ parents are designated by numbers representing the extra chromosome. Percentage of parental types in offspring shown by heights of vertical bars.

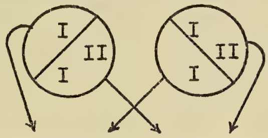
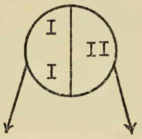
If the primaries had had viabilities equal to normals they would have constituted 50 per cent of the offspring from primary parents. To raise the number of primaries to this proportion, a viability correction (V. C.) is necessary in each case. This may be obtained by dividing the number of non-primaries in the offspring by the number of primaries. Thus the V. C. for $2n + 1 \cdot 2$ is 89.49 ÷ 10.51, or 8.51. The calculated V. C. for each primary is given in table 10. We are indebted to Dr. C. B. Bridges for assistance in working out formulae involved in these and other calculations in the present study. With a viability correction for primaries available it is possible to calculate the percentages of the different types of disjunction which take place at meiosis in secondaries. The results are shown in table 4. The disjunctions which give rise to $1n$ daughter cells should also give rise to equal numbers of $1n + \text{II}$ daughter cells. For example, in the case of the $2n + 1 \cdot 1$ secondary we can write down for both the $1n$ and the $1n + \text{II}$ gametes the observed percentage of $2n$ offspring (counting unrelated types as $2n$). The observed percentage of its primary $2n + 1 \cdot 2$ thrown by $2n + 1 \cdot 1$ parents may be multiplied by the V. C. for $1 \cdot 2$, namely 8.51, to give a product of 20.42, which represents the proportional number of this primary which would have been in the offspring if its viability had been equal to that of normals. The number of the daughter cells which lack a $1 \cdot 2$ chromosome but which have an extra $1 \cdot 1$ chromosome would also equal 20.42. From the total of these four figures may be recalculated the percentages of the different types of daughter cells. These are given in table 4.

From table 3 it was seen that there is considerable difference in the proportion of primaries thrown by the different secondaries, from 0.07 per cent for the secondary $2n + 19 \cdot 19$ to 8.46 per cent for $2n + 9 \cdot 9$. Table 4 shows that these differences are not due alone to differences in viability. The disjunction responsible for $1n + \text{I}$ daughter cells is lowest for the secondary $2n + 19 \cdot 19$ but highest for $2n + 2 \cdot 2$. In comparison with the latter, the disjunction giving $2n + \text{I}$ daughter cells comes out somewhat lower for $2n + 9 \cdot 9$ on account of the relatively high viability of its $2n + 9 \cdot 10$ primary.

The disjunctions which give $1n + \text{I}$ daughter cells should produce an equal number of cells deficient for the part of the chromosome represented in double dose in the complementary secondary. Pollen grains with such deficiencies should abort. In an earlier publication (25) are given the results of a study of the percentage of pollen abortion for the 12 primaries and for 10 of the secondaries. It is probably significant that the 5 secondaries with the highest percentages of disjunctions which give deficiencies (each over 8 per cent) showed an average of 13.8 per cent aborted pollen grains, while the 5 with the lowest percentage (each under 8 per cent) showed an average of 8.2 per cent aborted pollen. A close correlation is not expected since there is evidence that $1n + \text{II}$ pollen grains may also abort due to chromosomal unbalance although no deficiencies are present.

If there were random assortment of chromosomes in the trisome of secondaries one would expect disjunctions giving 33.33 per cent each of 1*n* and 1*n* + II daughter cells and 16.67 per cent each of 1*n* + I daughter cells and of daughter cells with deficiencies.

TABLE 4. PRODUCTS OF MEIOTIC DISJUNCTION IN 2*n* + 1 SECONDARIES (II)
(Calculated percentages)

Disjunction in trisomic set				
Parents	1*n*	1*n* + II	1*n* + I	1*n* − I + II
2*n* +				
1·1	41.17	41.17	8.83	8.83
2·2	34.58	34.58	15.42	15.42
3·3	43.65	43.65	6.35	6.35
5·5	42.04	42.04	7.96	7.96
6·6	42.51	42.51	7.49	7.49
7·7	40.44	40.44	9.56	9.56
9·9	40.36	40.36	9.64	9.64
10·10	46.50	46.50	3.50	3.50
11·11	47.55	47.55	2.45	2.45
13·13	42.25	42.25	7.75	7.75
14·14	45.14	45.14	4.86	4.86
15·15	42.36	42.36	7.64	7.64
17·17	40.76	40.76	9.24	9.24
19·19	48.43	48.43	1.57	1.57
Random assortment	33.33	33.33	16.67	16.67

These proportions are closely approached only in the case of the secondary 2*n* + 2·2. It is clear that the other disjunctions are not at random and that the different chromosomes have different effects upon the method of segregation. In the foregoing discussion we have spoken as if the three chromosomes in the trisome of a secondary were attached to each other at first metaphase. This is not always the case. More frequently the two normal chromosomes pair by themselves and separate to opposite poles while the II chromosome becomes attached to itself by its like ends to form a "doughnut" which lags and then goes at random to one pole or the other (20). This type of disjunction would tend to decrease the proportion of daughter cells with $1n + \text{I}$ and $1n - \text{I} + \text{II}$ chromosomes. The influence of different chromosomes upon the products of disjunction might be exercised by the frequency with which the II chromosome paired with itself to form a "doughnut." On this assumption "doughnut" formation would be relatively rare in $2n + 2\cdot2$ and almost the exclusive procedure in $2n + 19\cdot19$.

TABLE 5. OFFSPRING OF TETRASOMIC $2n + (15\cdot16)_2$

(Percentages in parentheses)

Parents	Plants potted	Dead and indeterminate	Recordable	$2n$	$2n + 15\cdot16$	$2n + (15\cdot16)_2$	Other types
$2n \times 2n + (15\cdot16)_2$	327	19	308	90 (29.22)	217 (70.45)	—	$2n + 17\cdot18$
$2n + (15\cdot16)_2 \times 2n$	71	10	61	22 (36.07)	39 (63.93)	—	
$2n + (15\cdot16)_2$ selfed	72	3	69	19 (27.54)	34 (49.28)	15 (21.74)	$2n + 15\cdot15$

Evidence regarding differences in metaphase behavior of these two contrasting secondaries is, however, lacking.

As shown in table 2, simple tetrasomic plants have been identified for 3 of the 4 primary chromosomes which can be transmitted as extras through the pollen. Records for selfs and for male and female back-crosses of the tetrasomic $2n + (15\cdot16)_2$ are shown in table 5. The germination of plantings from selfs and female back-crosses was rather poor and consequently in each case the best culture was chosen for the table. The male back-cross tabulations were taken from several crosses to gene types. The addition of extra chromosomes which when homozygous should produce new pure-breeding types with a new pair of chromosomes has been suggested by some as a method of evolution in genera having a series of species in which the number of pairs of chromosomes increases arithmetically. Records in our tetrasomics offer no support for such an hypothesis. Although there were four chromosomes in the 15·16 set, there was a considerable proportion of $2n$ individuals but no $2n + (15 \cdot 16)_2$ individuals in the back-cross offspring. The disjunctions of the tetrasome which gave rise to $1n$ gametes should also give rise to $1n + (15 \cdot 16)_2$ daughter cells. The failure of $2n + (15 \cdot 16)_2$ types to appear in back-cross offspring would indicate that such $1n + (15 \cdot 16)_2$ cells were inviable in the male and female gametophytes. The self, with 22 per cent $2n + (15 \cdot 16)_2$ offspring, runs lower in tetrasomics than would be expected from a combination of the gametes indicated by the male and female back-crosses and the $2n$ individuals run somewhat higher. The deficits may be due to selection in the zygote against two extra chromosomes.

TABLE 6. OFFSPRING OF FRAGMENT TYPE $2n + \cdot 11$ AND OF $2n + (\cdot 11)_2$
(Percentages in parentheses)

Parents	Plants potted	Dead and indeterminate	Recordable	$2n$	$2n + \cdot 11$	$2n + (\cdot 11)_2$
$2n + \cdot 11 \times 2n$	80	6	74	45 (60.81)	29 (39.19)	—
$2n + \cdot 11$ selfed	181	38	143	69 (48.25)	63 (44.06)	11 (7.69)
$2n + (\cdot 11)_2 \times 2n$	68	4	64	20 (31.25)	44 (68.75)	—
$2n + (\cdot 11)_2$ selfed	102	24	78	2 (2.56)	29* (37.18)	47 (60.26)

* Includes one $2n + 11 \cdot 12 + \cdot 11$.

The type with an extra ·11 fragment which is listed in table 2 has been selfed and female back-crossed with results shown in table 6. Male back-crosses were not made but from the selfs it is evident that the fragment can function when carried by the pollen. The progeny from selfing a parent heterozygous for the fragment contains somewhat more $2n$ and somewhat fewer $2n + (\cdot 11)_2$ individuals than one might expect from the back-cross ratio. The plant homozygous for the fragment when selfed does not breed true. Although there is a high proportion of $2n + (\cdot 11)_2$

individuals in the progeny, there are a considerable number of $2n + \cdot 11$ types and a few normals ($2n$). The female back-cross of the plant homozygous for the fragment throws over 30 per cent $2n$ offspring. The equal number of $2n + (\cdot 11)_2$ offspring which might have been expected were probably prevented from appearing by the resistance of the gametophyte to so great an unbalance as two extra fragments. A number of other fragment types which we have investigated show a breeding behavior essentially similar to the one just discussed. Some of the fragments cannot be transmitted through the pollen and from these we have never secured homozygous individuals.

TABLE 7. OFFSPRING OF DOUBLE TRISOMIC ($2n + 1 + 1$) TYPES

Parents	Number of parents	Plants potted	Dead and indeterminate	Recordable	$2n$	Parent type $2n$ + Ia + Ib	$2n$ + Ia	$2n$ + Ib	Other types
$2n$ +									
1·2 + 9·10	2	74	12	62	37	—	6	19	
1·2 + 23·24	1	36	—	36	22	1	2	10	$2n$ + 2·2
3·4 + 11·12	1	26	1	25	21	—	2	2	
5·6 + 9·10	3	316	21	295	160	25	58	51	$2n$ + 5·6 + 13·14* $2n$ + 23·24
5·6 + 11·12	1	89	11	78	29	4	16	29	
5·6 + 17·18	1	95	7	88	62	1	10	15	
5·6 + 21·22	1	126	4	122	74	4	25	19	
5·6 + 23·24	1	55	1	54	14	7	16	17	
13·14 + 15·16	1	20	1	19	10	—	3	6	
15·16 + 23·24	3	414	17	397	212	22	81	82	
Totals............	15	1251	75	1176	641	64	219	250	
Percentages........					54.51	5.44	39.87		

* Included also in $2n$ + Ia.

In table 7 are given data on the offspring from selfing some of the double trisomics ($2n + 1 + 1$). If viability were equal for all the types in the offspring one might expect progenies consisting of (1) $2n$: (1) $2n$ + Ia + Ib : (1) $2n$ + Ia : (1) $2n$ + Ib, when Ia and Ib represent the two primary chromosomes present as extras in the double trisomic. Of the three double trisomics which have more than 100 offspring recorded, the $2n + 5\cdot 6 + 9\cdot 10$ gives a ratio of the above types of roughly 6:1:2:2; $2n + 5\cdot 6 + 21\cdot 22$

gives a ratio of 18:1:5:5 and $2n + 15\cdot16 + 23\cdot24$ gives a ratio of 10:1:4:4. It will be noticed that the double and simple trisomics have greatly lessened viability in comparison with $2n$ offspring, the double types naturally suffering most from the unbalance. The proportion of Ia and Ib types depends upon their relative viability. When the Ia type has the $1\cdot2$ chromosome as an extra, the viability is greatly decreased as shown by the low numbers of the simple Ia trisomic as well as of the double type with the Ia chromosome as one of the extras. The final percentages for this table are calculated from the total numbers and are loaded unduly by the larger cultures. Some of the cultures, however, appeared to have too small numbers to justify individual percentages.

TABLE 8. OFFSPRING OF TERTIARY (III) $2n + 1$ TYPES

Parents	Per cent $2n + 1$ offspring				Per cent $2n$ offspring	Totals
	Parent type	Related types		Unrelated types		
		I	II			
$2n +$						
$1\cdot9$	14.83	1.57	0.12	0.16	83.31	2487
$2\cdot5$	28.91	1.72	0.05	0.81	68.74	2089
$4\cdot6$	30.07	0.16	0.04	0.0	69.73	2458
$1\cdot18$	16.09	0.73	0.0	0.17	83.02	1790
$2\cdot17$	20.57	0.53	0.04	0.04	78.82	2460
Averages....	22.09	0.94	0.05	0.24	76.72	

In table 8 are summarized in percentages the data on inheritance of tertiary $2n + 1$ types from table 2. From these two tables it will be seen that tertiaries throw more related primaries than they do any other trisomics. The tertiary $2n + 4\cdot6$ was earlier thought (29) to be related only to the primary $2n + 5\cdot6$ since it did not appear to throw an adequate proportion of $2n + 3\cdot4$ primaries to show relationship to the latter. Additional breeding records from $2n + 4\cdot6$ parents heterozygous for various lines have given, in over 2000 offspring, the following primaries: (11) $2n + 5\cdot6$, (1) $2n + 3\cdot4$, (1) $2n + 11\cdot12$, (2) $2n + 13\cdot14$, (1) $2n + 17\cdot18$ and (1) $2n + 4\cdot6 + 21\cdot22$. Later, linkage with genes as well as cytological study proved that $2n + 4\cdot6$ is entitled to the formula which we have given it. The production of related primaries from tertiaries is a matter of segregation due to the type of disjunction of chromosomes connected at metaphase.

Trisomics as Mutation Producing Machines

In the previous section on inheritance of extra chromosomes we have discussed the appearance of the same and of related types in the offspring of $2n + 1$ parents. The transmission of extra chromosomes has been related to the way in which the chromosomes segregate at meiotic divisions. In the present section will be discussed the spontaneous appearance in the offspring of extra chromosomal types which are unrelated to the parental type. Their origin may be classed as chromosomal mutation. Table 9 is a

TABLE 9. PERCENTAGE OF UNRELATED TYPES THROWN BY EACH $2n + 1$ PRIMARY AND SECONDARY

Parents	Per cent new mutants		Rank of parents in producing mutants	
	From I	From II	I	II
$2n$ +				
1·1		0.20		13
1·2	1.33		3	
2·2		0.38		10
3·3		0.79		5
3·4	0.05		12	
5·5		0.98		3
5·6	1.32		4	
6·6		0.13		14
7·7		0.22		12
7·8	0.29		9	
9·9		1.34		2
9·10	0.74		7	
10·10		1.45		1
11·11		0.49		8
11·12	0.76		6	
13·13		0.39		9
13·14	1.63		2	
14·14		0.80		4
15·15		0.63		6
15·16	0.97		5	
17·17		0.31		11
17·18	0.56		8	
19·19		0.59		7
19·20	2.27		1	
21·22	0.26	—	10	—
23·24	0.10	—	11	—
Averages......	0.86	0.62		

percentage summary from table 2. The data are shown in graphs in figure 2. We are here discussing the differences between the various primary and secondary trisomics as mutation producing machines. Of the primaries, $2n + 19 \cdot 20$ is outstanding in stimulating new mutations (2.27 per cent) while $2n + 3 \cdot 4$ is the poorest as a mutation producing machine, giving an output of only .05 per cent new mutants. This rate is considerably less than that from $2n$ parents (.16 per cent). There seems no close connection between the size of a chromosome and its ability to produce new mutations. The primaries, however, seem better machines than the secondaries. This difference can be accounted for to only a slight extent by the fact that secondaries regularly throw related primaries which would have considerably lessened viability if combined with new mutant trisomics to form double $2n + 1 + 1$ types. In only two cases ($2n + 3 \cdot 4$ and $2n + 9 \cdot 10$) does the primary fail to be a better mutation machine than one or both of its secondaries. In their influence upon the spontaneous appearance of new mutants the primaries are not intermediate between their two secondaries as they tend to be in respect to morphological characters (54). Primaries throw an average of .86 per cent new mutants; secondaries an average of .62 per cent. Both form better mutation machines than do $2n$ parents which throw .16 per cent

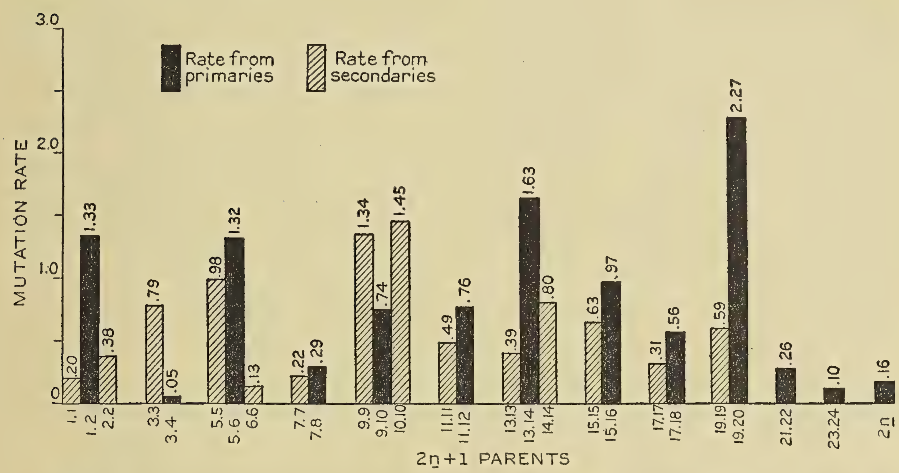


FIG. 2. Mutation rates from primary and secondary $2n + 1$ parents. Vertical bars show mutation rates in terms of percentages of unrelated types thrown by each $2n + 1$ parent.

new trisomics. Diploids can produce unrelated mutations in all 12 chromosome sets, primaries and secondaries, however, only in 11. Why the presence of one extra chromosome increases the chromosomal mutation rate is not clear. The increase is perhaps in some way caused by interference with meiotic divisions. It cannot, however, be directly related to size of chromosomes as will be seen from the graphs and tables. Moreover, the next to the largest chromosome (3·4) may have even an inhibiting effect upon the appearance of spontaneous chromosomal mutations if the rate of .05 per cent from 2000 offspring is statistically significant. This rate is a third of that from $2n$ parents.

From table 8, it will be seen that the mutation rate from the tabulated tertiaries (averaging 0.24 per cent) is relatively low. It is known in some cases that the breaks at the formation of tertiaries are at a different level from those at the formation of secondaries. Moreover, it will be seen from the graphs of figure 2 that there is little relation apparent between the effects of a pair of secondary chromosomes and the primary chromosome which is compounded of parts of these same secondaries. For these reasons as well as because of the few tertiaries represented it will not be possible to draw conclusions regarding any relationships between the effects of tertiary chromosomes and the parts of chromosomes of which they are composed.

Since the presence of a single extra chromosome in the parent has been found to increase the new occurrence of unrelated $2n + 1$ types in the offspring, it would be reasonable to expect that two extra chromosomes would still further increase the mutation rate. There is evidence rather that the addition of a second extra chromosome actually decreases the mutation rate. From table 7 it will be seen that there were only two unrelated non-disjunctional types in 1176 offspring from double trisomic ($2n + 1 + 1$) parents. From the mutation rates from parents with single extra chromosomes given in table 9, it is possible to calculate how many new mutations might be expected if only one extra chromosome were present in each parent. The calculations take account of the fact that there can be only 10 unrelated primaries in the offspring of a $2n + \mathrm{I} + \mathrm{I}$ parent. Without correction for lessened viability of $2n + 1 + 1$ types and for the practical non-existence of triple trisomics ($2n + 1 + 1 + 1$) under the existing conditions of culti vation, the quota of new mutants to be expected among the 1176 offspring would be more than a dozen rather than the 2 actually found. It is difficult to make the suggested correction for viability of $2n + 1 + 1$ types but the decrease of new mutants in the offspring when the parents have two extra chromosomes instead of only one may be shown in another way. The plus 1 types in the progeny can be disregarded entirely and a quota of new mutants can be calculated based on the number of $2n$ offspring alone. Although this calculation gives an expected mutation rate greatly below that which actually exists, it shows that more than 7 new mutants should have appeared if each parent had had only a single extra chromosome, whereas with two extra chromosomes the number actually observed is 2. The reason why a second extra chromosome decreases the mutation rate is not clear.

There are not sufficient numbers in the offspring of tetrasomics in table 5 to give statistically significant rates of mutation. However, so far as they go they offer no support for an increase of mutation rate due to a second extra chromosome in the same set. The $2n + 17\cdot18$ mutant from the cross $2n \times 2n + (15\cdot16)_2$ must have come from the $2n$ parent since the $17\cdot18$ chromosome of Line 1 appears not to be transmissible through the pollen. With the mutation rate of .97 from simple trisomic $2n + 15\cdot16$ female parents the quota for the 130 offspring in the table should have given more than 1 unrelated mutant type if only a single chromosome were extra. Since the $2n + 15\cdot15$ secondary is related to the parent type we have no observed mutant types from tetrasomic female parents.

Records have been tabulated from too few tertiaries to show whether tertiaries as such are better mutation machines than primaries and secondaries. The mutation rate from the tertiaries shown in table 8 (0.24) averages higher than that from $2n$ parents (0.16) but lower than the average from primaries (0.86) or from secondaries (0.62). The mutation rate from the $2n + 2\cdot5$ tertiary is higher than that from $2n$ parents while the rates from $2n + 2\cdot17$ and from $2n + 4\cdot6$ parents are lower.

$2n + 1$ Types as Products of Mutation

In the previous section we have discussed the efficiency of the $2n + 1$ types as machines for the production of unrelated $2n + 1$

mutants. It was seen that they differed greatly in their ability to stimulate new chromosomal mutations. In the present section we shall present data to show which chromosomes are more likely to be used as extras in the formation of spontaneous $2n + 1$ types. We are now considering the new $2n + 1$ types as products of chromosomal mutation. The raw data are given in table 2. Table 10 gives the percentage that each new $2n + 1$ mutant type forms of the offspring (1) of diploids, (2) of unrelated primaries, and (3) of unrelated secondaries. The origin of a $1n + 1$ gamete presumably traces back to non-disjunction of a single disome, leaving an $n - 1$ daughter cell which would not be able to survive in the gametophyte. Due to the fact that the 12 primary chromosomes differ in viability, it will not be possible to use without correction the percentage of a new $2n + 1$ type in the offspring as an index of the rate at which non-disjunction of the primary chromosome takes place. The viability correction (V. C.) which has already been explained is given in the second column of table 10. If M

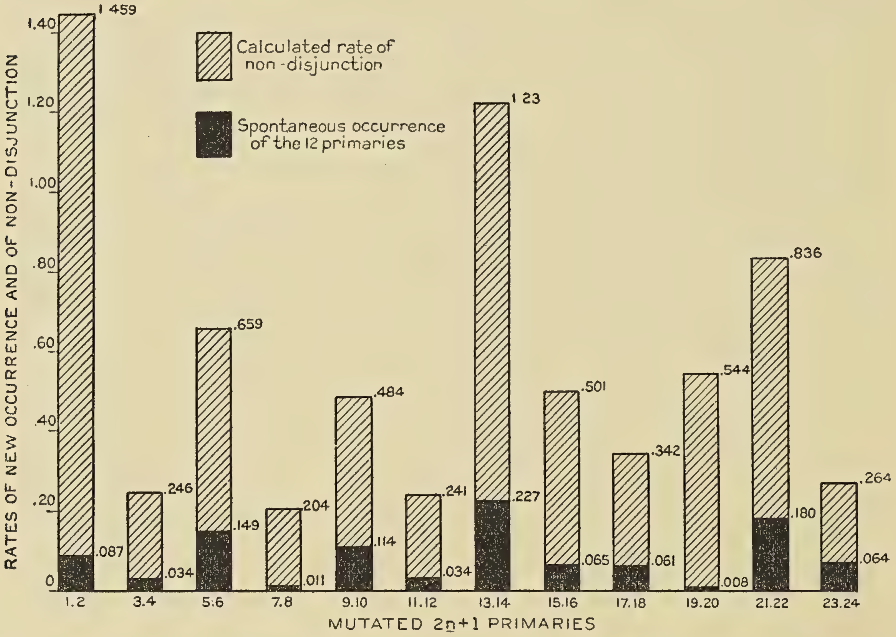


Fig. 3. Rate of occurrence of the 12 primary $2n + 1$ types in offspring from unrelated primary parents and calculated rate of non-disjunction responsible for such new mutant types.

represents the observed spontaneous occurrences of a new mutant of a given primary type and N the normals (including as normals all individuals except the primary mutants in question), then the rate of non-disjunction will be the corrected mutant rate divided by the corrected totals or $\frac{2M \times VC}{N + 2M \times VC}$. It will be observed that the rank in non-disjunction for a given chromosome may be quite different from its rank in observed number of mutants. Thus on account of poor viability the calculated percentage of non-disjunction responsible for the $2n + 19 \cdot 20$ mutant types is 60 times the percentage of new mutants of this type which have been observed in our cultures. Figure 3 presents graphs which show in percentages the observed spontaneous occurrence of the different primaries together with the calculated percentages of non-disjunction responsible for their origin. The figures are taken from the mutation rates from unrelated primaries given in the middle section of table 10.

TABLE 10. MUTATION RATE OF THE 12 PRIMARIES

Percentages of observed new occurrences of the 12 primary $2n + 1$ types and the calculated non-disjunction responsible for their origin.

Primary mutants	Viability correction	From Diploids			From Primaries			From Unrelated Secondaries		
		Observed	Calculated non-disjunction	Rank	Observed	Calculated non-disjunction	Rank	Observed	Calculated non-disjunction	Rank
$2n +$										
1·2	8.51	0.030	0.503	1	0.087	1.459	1	0.111	1.857	1
3·4	3.62	—	—		0.034	0.246	10	0.021	0.148	10
5·6	2.22	0.015	0.066	8	0.149	0.659	4	0.099	0.440	3
7·8	8.99	—	—		0.011	0.204	12	0.004	0.062	12
9·10	2.13	0.020	0.084	6	0.114	0.484	7	0.058	0.248	7
11·12	3.52	—	—		0.034	0.241	11	0.014	0.096	11
13·14	2.73	0.030	0.162	3	0.227	1.226	2	0.132	0.716	2
15·16	3.88	0.020	0.153	4	0.065	0.501	6	0.044	0.341	5
17·18	2.82	0.020	0.112	5	0.061	0.342	8	0.044	0.248	6
19·20	32.47	0.005	0.320	2	0.008	0.544	5	0.004	0.225	8
21·22	2.34	0.015	0.069	7	0.180	0.836	3	0.088	0.412	4
23·24	2.06	—	—		0.064	0.264	9	0.044	0.182	9
Averages		0.013	0.122		0.086	0.584		0.055	0.414	

There does not seem to be any relation between size of chromosome and its likelihood of taking part in non-disjunction. The disome most frequently suffering non-disjunction contains the largest chromosome 1·2, the next most frequent contains a medium-sized chromosome 13·14, while the third most frequent contains a small chromosome 19·20. Of the four chromosomes with a non-disjunction rate below 0.27, 3 are among the largest 6 chromosomes.

The fact that the non-disjunction rate is greater from trisomics than from diploids indicates that the presence of one extra chromosome in some way stimulates non-disjunction in other sets. Only for the 1·2 chromosome is the non-disjunction rate greater from secondaries than from primaries. The average rate from secondaries is 0.414 against 0.584 from primaries. It is possible that disjunction from a trisome interferes with orderly meiosis and favors non-disjunction. The fact that frequently the secondary chromosome at metaphase forms a rounded "doughnut" unconnected with its related primary chromosomes which pair normally may be responsible for the lessened non-disjunction from secondaries.

Whereas all the primary $2n + 1$ types have appeared repeatedly in our cultures as spontaneous mutants, brought about by non-disjunction, they arise in much higher numbers through segregation in the offspring of triploids (29). An exception is the primary $2n + 19 \cdot 20$, the viability of which is poor and which appears rarely also as a spontaneous mutant. Our common procedure in getting the primary $2n + 1$ types heterozygous for a gene which we wish to locate in the proper chromosome by trisomic ratios is to male-cross the gene to triploids. Out of less than 150 offspring we are almost certain to secure all the 12 primaries except $2n + 19 \cdot 20$. The double trisomic primaries ($2n + 1 + 1$) also arise through segregation from triploids and appear in the offspring of $2n + 1$ parents. In the latter case one of the extra chromosomes is due to segregation and the other to non-disjunction. An example from table 2 is the appearance of a $2n + 5 \cdot 6 + 11 \cdot 12$ from a $2n + 5 \cdot 6$ parent. Rarely simultaneous non-disjunctions in two sets give rise to a double trisomic unrelated to the parental type. An example is $2n + 11 \cdot 12 + 21 \cdot 22$ from the $2n + 13 \cdot 14$ parent. There are theoretically possible 66 double trisomics each with a different combination of 2 extra primary chromosomes. Due to the decreased viability of some of these types we have not found all of the 66, but 52 different double trisomic types have been identified. Capsules of some of these have been figured in an earlier publication (43).

The origin of secondary $2n + 1$ types presents a special problem. From tables 2, 11, and 12, it will be seen that secondaries arise spontaneously from unrelated primary parents (0.03 per cent) and rarely from unrelated secondaries (0.001 per cent) and $2n$ parents. They arise in offspring of a related primary (0.07 per cent) more than twice as often as from an unrelated primary. They have not been observed in the offspring from related secondaries but do appear in the offspring of related tetrasomics (table 5) and

TABLE 11. SPONTANEOUS OCCURRENCE OF SECONDARY $2n + 1$ TYPES

Secondary types	Secondaries from $2n$ parents	Secondaries from $2n + 1$ parents			
	Totals (19820 offspring)	From related I's	From unrelated I's	From unrelated II's	Totals (60191 offspring)
$2n +$					
1·1	—	1	1	—	2
2·2	—	5	1	1	7
3·3	—	—	1	1	2
5·5	—	2	1	1	4
6·6	—	—	1	—	1
7·7	—	—	—	—	—
9·9	—	1	1	—	2
10·10	—	—	—	—	—
11·11	—	2	1	—	3
13·13	—	—	1	—	1
14·14	—	1	—	1	2
15·15	—	1	1	—	2
17·17	—	1	—	2	3
19·19	1	7	—	—	7
Totals........	1	21	9	6	36
Percentages...	0.005	0.07	0.03	0.001	0.06

of related tertiaries (tables 2 and 8). The percentage of secondaries thrown depends upon which primary is the parent and also upon the viability of the secondary. Direct evidence is lacking as to the exact mechanism involved in the origin of secondaries. Belling (20) suggested that secondary chromosomes might originate from crossing-over between two sister strands which lay side by side but with the ends of one reversed. This arrangement with the 1·2 chromosome for example would give the secondary 1·1 and 2·2 chromosomes. If the parts of the two strands were in reversed order there would be only the center of the strands where parts would be homologous and it was at this point at which the breaks were believed to occur. Another suggested method is unequal crossing-over between parallel sister strands in such a way as to retain spindle attachment points for each newly organized secondary chromosome which has been formed by joining together by their broken ends the two similar halves of the strands affected. How the secondary chromosome once formed could be gotten into a viable gamete is a problem probably involving non-disjunction. The frequency with which secondaries arise from their own primaries is greater probably because with an extra chromosome there are

TABLE 12. RELATED AND UNRELATED SECONDARIES THROWN BY PRIMARIES

Primary parents	Recordable	Related II's thrown		Unrelated II's thrown	
		Number	Per cent	Number	Per cent
$2n +$					
1·2	2026	6	0.29	0	0.0
3·4	2087	0	0.0	0	0.0
5.6	2349	2	0.09	3	0.13
7·8	2079	0	0.0	0	0.0
9·10	2157	1	0.05	0	0.0
11·12	2227	2	0.09	1	0.04
13·14	2025	1	0.05	1	0.05
15·16	2276	1	0.04	0	0.0
17·18	2136	1	0.05	0	0.0
19·20	4753	7	0.15	4	0.08
21·22	2329	—	—	0	0.0
23·24	2038	—	—	0	0.0
Totals........	28482	21	0.07	9	0.03

more opportunities for crossing-over and for the secondary chromosome to be gotten into a gamete through trisomic disjunction.

From table 11 it will be seen that most of the secondaries have arisen more than once in the Line 1 cultures under discussion. They have arisen also from other sources. From other sources also the $2n + 7 \cdot 7$ secondary has arisen twice, the $2n + 6 \cdot 6$ once, the $2n + 13 \cdot 13$ once. The $2n + 10 \cdot 10$ secondary is the only one which has arisen only once. It appeared in 1926 in the offspring of a $2n - 1$ type. The $2n + 1 \cdot 1$ secondary has appeared 6 times; the $2n + 2 \cdot 2$, thirteen times; the $2n + 5 \cdot 5$, 9 times; the $2n + 19 \cdot 19$, 11 times. Since no new secondaries have appeared in the last nine years, it is unlikely that many more will be discovered. Rarity of a secondary may be due to poor viability after the secondary is formed or to absence of the peculiar behavior of the primary chromosome necessary for its origin. There is no known method of inducing the production of secondaries.

It has not been possible to detect any differences in appearance between the different occurrences of a given secondary type. It is likely therefore that the breaks responsible for the origin of secondary chromosomes always occur at the same place or nearly at the same place.

The first tertiary $2n + 1$ recognized in our cultures was called Wiry (Wy). Its relationship to the $2n + 17 \cdot 18$ primary was early recognized by similarity in morphological peculiarities, by throwing $2n + 17 \cdot 18$ plants in its offspring and by showing trisomic ratios for genes located in the $2n + 17 \cdot 18$ chromosome. It was noted that it appeared only in cultures which were at least heterozygous for certain types of races from nature which from their breeding behavior were known as B races in distinction from the A races found in nature to which our Line 1 belongs. It was from the attachments of chromosomes in this Wiry that Belling (24) found the clue to the nature of the B races and was thereby enabled to propose his fruitful hypothesis of segmental interchange. Wiry is now known as the $2n + 1 \cdot 18$ tertiary and the other tertiary from the B race (now known as Prime Type 2) is $2n + 2 \cdot 17$. Prime Types have been discussed in an earlier paper (48). They are found in nature (68) and readily induced by radiation, aging seeds (59), and by heat (65). Most of the prime types in our collection are homozygous for chromosomes modified by segmental interchange. Prime Type 2 for example contains the chromosomes 1·18 and 2·17 in place of the chromosomes 1·2 and 17·18 of our Line 1. All the other tertiaries in table 8 have appeared in our cultures of Line 1. The two tertiaries $2n + 1\cdot 9$ and $2n + 2\cdot 5$ came from a $2n$ plant treated by Gager (26) with radium emanation in 1921. Together they form the Compensating Type called Nubbin (29). It apparently arose through interchange between the 1·2 and 9·10 chromosomes as well as between the 1·2 and 5·6 chromosomes.

The spontaneous occurrence of tertiaries has been rare in our cultures. The earliest one identified was $2n + 2\cdot 9$ discovered in 1923 in the offspring of $2n + 11\cdot 12$. It is the only spontaneous tertiary of ours which has come from an unrelated $2n + 1$ type. It has been found only once. The $2n + 4\cdot 6$ tertiary has appeared twice, once in 1921 in the offspring of its related primary $2n + 3\cdot 4$ and again in 1924 in the offspring of the related secondary $2n + 3\cdot 3$. This latter plant was tested cytologically and found not to contain a complementary 3·5 chromosome which might have been expected as a sister product of interchange between the 3·4 and 5·6 chromosomes presumably responsible for the origin of the 4·6 chromosome.

The 13·18 chromosome has appeared two times as a component of the compensating type $2n - 17\cdot 18 + 17\cdot 17 + 13\cdot 18$. It has not appeared spontaneously alone but has become separated out in a $2n + 13\cdot 18$ tertiary in the offspring of the compensating type. This compensating type first appeared in the offspring of the related $2n + 17\cdot 17$ secondary. In 1928 the same compensating type appeared spontaneously from a parent heterozygous for a fragment of the 17·18 chromosome. In this latter case both the 17·17 and the 13·18 chromosomes must have been formed *de novo* since neither was in the formula of its parent. One of the compensating types was tested cytologically and found not to contain the other complementary 14·17 chromosome.

The occurrence in nature of prime types with interchanged chromosomes, as well as the spontaneous appearance in our cultures of the tertiaries just mentioned led us to believe that segmental interchange might be taking place and establishing prime types in our cultures without our realizing it. At this time we had all our primaries and present secondaries which had been kept going by selfing continuously, some of them for as many as 10 generations. If segmental interchange were common, some of these inbred lines might have established themselves as prime types with modified chromosomes such as those in the tertiaries which had appeared spontaneously. To test this possibility Dr. Bergner studied cytologically F_1 hybrids between our Line 1 and 24 of the 26 primary and secondary inbred races but found no evidence that segmental interchange leading to prime types had taken place. Although new prime types are readily induced by various treatments, their establishment in nature seems to be relatively rare. From study of over 600 races from localities throughout the world (68), only 5 prime types have been found to be common and of wide distribution. There were 4 other prime types which were apparently sporadic, each having been found only once. With the paucity of prime types in nature and the rarity of secondaries and tertiaries in our own cultures it may be said that the individual chromosomes of *Datura stramonium* appear to be relatively stable in constitution.

The 15-year breeding records of extra chromosomal types in *Datura* emphasize the individuality of the chromosomes and the influence they individually exert upon the viability of the plant and its constituent cells as well as upon chromosome behavior which may lead to the production of specific types of offspring differing in number and arrangement of chromosomes.

SUMMARY

Records on over 60,000 field-grown offspring of 12 primary and 14 secondary $2n + 1$ types lead to the following conclusions:

1. Both primaries and secondaries differ widely in viability (transmission of the extra chromosome to their offspring). The highest viability is shown by $2n + 23 \cdot 24$ with 32.7 per cent transmission and the lowest by $2n + 19 \cdot 20$ with 3.0 per cent transmission.
2. The number of primaries thrown by their own secondaries also varies widely even when percentages are corrected for viability. This fact indicates that disjunctions responsible for $1n + 1$ primary gametes are not at random.
3. The number of secondaries thrown by unrelated primaries is

extremely small (0.03 per cent). The number of secondaries thrown by their own primaries is slightly larger (0.07 per cent).

4. There is a marked difference in respect to the number of new mutations of unrelated $2n + 1$ types which each type throws (0.05 per cent from the $2n + 3\cdot4$ type to 2.3 per cent from the $2n + 19\cdot20$ type). Primaries and secondaries together throw 0.7 per cent unrelated $2n + 1$ types (0.9 per cent from primaries and 0.6 per cent from secondaries). Since among nearly 20,000 offspring from diploids there were only 0.16 per cent $2n + 1$ types, it is evident that non-disjunction must be more frequent in $2n + 1$ types than in diploids. A second extra chromosome in a double trisomic ($2n + 1 + 1$) appears to decrease the mutation rate of unrelated $2n + 1$ types.

5. Simple tetrasomic types ($2n + 2$) do not breed true. In their offspring from selfing are found $2n + 2$, $2n + 1$ and $2n$ individuals. In back-crosses $2n + 2$ plants fail to appear.

6. Tertiary $2n + 1$ types (the extra chromosome of which is composed of parts of two normal chromosomes) give in their offspring diploids, tertiaries, and their two related primaries.

7. Certain types are more frequent among the new mutations than others even when figures are corrected for differences in viability. The conclusion is reached that chromosomes differ in their frequency of primary non-disjunction. These differences are not related to chromosome size.

BIBLIOGRAPHY

1920

1. Blakeslee, A. F., John Belling, and M. E. Farnham, Chromosomal Duplication and Mendelian Phenomena in *Datura* Mutants, *Science*, n.s., vol. 52, no. 1347, pp. 388–390 (Oct.).

1921

2. ——— A Graft-infectious Disease of *Datura* Resembling a Vegetative Mutation, *Jour. Gen.*, vol. 11, pp. 17–36, pls. II–VI, figs. 1–12 (Apr.).
3. ——— The Globe Mutant in the Jimson Weed (*Datura stramonium*), *Genetics*, vol. 6, pp. 241–264 (May).
4. ——— Types of Mutations and Their Possible Significance in Evolution, *Amer. Naturalist*, vol. 55, pp. 254–267 (May–June).
5. ——— An Apparent Case of Non-Mendelian Inheritance in *Datura* Due to a Disease, *Proc. Nat. Acad. Sci.*, vol. 7, pp. 116–118 (Apr.).

6. Blakeslee, A. F., The Globe, a Simple Trisomic Mutant in *Datura*, *Proc. Nat. Acad. Sci.*, vol. 7, pp. 148–152 (May).

1922

7. Sinnott, E. W., and A. F. Blakeslee, Structural Changes Associated with Factor Mutations and with Chromosome Mutations in *Datura*, *Proc. Nat. Acad. Sci.*, vol. 8, pp. 17–19 (Feb.).
8. Blakeslee, A. F., Variations in *Datura* Due to Changes in Chromosome Number, *Amer. Naturalist*, vol. 56, pp. 16–31 (Jan.–Feb.).
9. Buchholz, J. T., and A. F. Blakeslee, Studies of the Pollen Tubes and Abortive Ovules of the Globe Mutant of *Datura*, *Science*, n.s., vol. 55, no. 1431, pp. 597–599 (June).
10. Blakeslee, A. F., John Belling, M. E. Farnham, and A. Dorothy Bergner, A Haploid Mutant in the Jimson Weed, *Datura stramonium*, *Science*, n.s., vol. 55, no. 1433, pp. 646–647 (June).
11. Belling, John, and A. F. Blakeslee, The Assortment of Chromosomes in Triploid *Daturas*, *Amer. Naturalist*, vol. 56, pp. 339–346 (July–Aug.).
12. Blakeslee, A. F., John Belling, and J. Arthur Harris, The Probability Established by a Culture of Given Size That a Mating is Capable of Producing only Dominant Individuals, *Amer. Naturalist*, vol. 56, pp. 458–461 (Sept.–Oct.).

1923

13. ——— Variations in the Jimson Weed (*Datura stramonium*) Caused by Differences in the Number of Chromosomes, *Eugenics, Genetics, and the Family*, vol. 1, pp. 82–83.
14. Belling, John, and A. F. Blakeslee, The Reduction Division in Haploid, Diploid, Triploid, and Tetraploid *Daturas*, *Proc. Nat. Acad. Sci.*, vol. 9, pp. 106–111 (Apr.).
15. Blakeslee, A. F., and M. E. Farnham, Bottle Grafting, *Jour. Hered.*, vol. 14, no. 4, pp. 171–173 (July).
16. ——— John Belling, and M. E. Farnham, Inheritance in Tetraploid *Daturas*, *Bot. Gaz.*, vol. 76, no. 4, pp. 329–373 (Dec.).
17. ——— and M. E. Farnham, Trisomic Inheritance in the Poinsettia Mutant of *Datura*, *Amer. Naturalist*, vol. 57, pp. 481–495 (Nov.–Dec.).

1924

18. Belling, John, and A. F. Blakeslee, The Distribution of Chromosomes in Tetraploid *Daturas*, *Amer. Naturalist*, vol. 58, no. 654, pp. 60–70 (Jan.–Feb.).
19. Blakeslee, A. F., Distinction between Primary and Secondary Chromosomal Mutants in *Datura*, *Proc. Nat. Acad. Sci.*, vol. 10, no. 3, pp. 109–116 (Mar.).
20. Belling, John, and A. F. Blakeslee, The Configurations and Sizes of the Chromosomes in the Trivalents of 25-Chromosome *Daturas*, *Proc. Nat. Acad. Sci.*, vol. 10, no. 3, pp. 116–120 (Mar.).

21. Blakeslee, A. F., Bottle Grafting Again, *Jour. Hered.*, vol. 15, p. 54 (Feb.).
22. ——— and John Belling, Chromosomal Chimeras in the Jimson Weed, *Science*, n.s., vol. 60, no. 1540, pp. 19–20 (July).
23. ——— ——— Chromosomal Mutations in the Jimson Weed, *Datura stramonium*, *Jour. Hered.*, vol. 15, no. 5, pp. 195–206 (May).

1926

24. Belling, John, and A. F. Blakeslee, On the Attachment of Non-Homologous Chromosomes at the Reduction Division in Certain 25-Chromosome *Daturas*, *Proc. Nat. Acad. Sci.*, vol. 12, pp. 7–11 (Jan.).
25. Blakeslee, A. F., and J. L. Cartledge, Pollen Abortion in Chromosomal Types of *Datura*, *Proc. Nat. Acad. Sci.*, vol. 12, pp. 315–323 (May).

1927

26. Gager, C. S., and A. F. Blakeslee, Chromosome and Gene Mutations in *Datura* Following Exposure to Radium Rays, *Proc. Nat. Acad. Sci.*, vol. 13, no. 2, pp. 75–79 (Feb.).
27. Blakeslee, A. F., The Chromosomal Constitution of Nubbin, a Compound ($2n + 1$) Type in *Datura*, *Proc. Nat. Acad. Sci.*, vol. 13, no. 2, pp. 79–85 (Feb.).
28. Buchholz, J. T., and A. F. Blakeslee, Abnormalities in Pollen-Tube Growth in *Datura* Due to the Gene "tricarpel," *Proc. Nat. Acad. Sci.*, vol. 13, no. 4, pp. 242–249 (Apr.).
29. Blakeslee, A. F., Nubbin, a Compound Chromosomal Type in *Datura*, *Annals New York Acad. Sci.*, vol. 30, pp. 1–29, pls. 1–8 (July).
30. ——— G. Morrison, and A. G. Avery, Mutations in a Haploid *Datura*, *Jour. Hered.*, vol. 18, no. 5, pp. 193–199 (May).
31. Buchholz, J. T., and A. F. Blakeslee, Pollen-Tube Growth at Various Temperatures, *Amer. Jour. Bot.*, vol. 14, no. 7, pp. 358–369 (July).
32. ——— ——— Pollen-Tube Behavior with Reference to Sterility in *Datura*, *Mem. Hort. Soc. New York*, vol. 3, pp. 245–260 (July).
33. Blakeslee, A. F., and J. L. Cartledge, Sterility of Pollen in *Datura*, *Mem. Hort. Soc. New York*, vol. 3, pp. 305–312 (July).
34. Belling, John, and A. F. Blakeslee, Assortment of Chromosomes in Haploid Daturas, *La Cellule*, vol. 37, pp. 353–365.

1928

35. Blakeslee, A. F., Genetics of *Datura*, *Verh. V. Internat. Kong. Vererb.*, vol. 1, pp. 117–130.

1929

36. ——— An Attempt to Analyse the Composition of Nubbin, a Compound ($2n + 1$) Chromosomal Type in *Datura*, *Proc. Internat. Cong. Plant Sciences*, vol. 1, pp. 831–832.
37. ——— Cryptic Types in *Datura* Due to Chromosomal Interchange, and Their Geographical Distribution, *Jour. Hered.*, vol. 20, no. 4, pp. 177–190 (Apr.).

38. Buchholz, J. T., and A. F. Blakeslee, Pollen-Tube Growth in Crosses between Balanced Chromosomal Types of *Datura stramonium*, *Genetics*, vol. 14, no. 6, pp. 538–568 (Nov.).

1930

39. Blakeslee, A. F., and R. E. Cleland, Circle Formation in *Datura* and *Oenothera*, *Proc. Nat. Acad. Sci.*, vol. 16, no. 3, pp. 177–183 (Mar.).

40. Buchholz, J. T., and A. F. Blakeslee, Pollen-Tube Growth of the Primary Mutant of *Datura*, Rolled, and Its Two Secondaries, *Proc. Nat. Acad. Sci.*, vol. 16, no. 3, pp. 190–195 (Mar.).

41. ——— ——— Radium Experiments with *Datura*, I: The Identification and Transmission of Lethals of Pollen-Tube Growth in F_1's from Radium-Treated Parents, *Jour. Hered.*, vol. 21, no. 3, pp. 119–129 (Mar.).

42. ——— ——— Pollen-Tube Growth and Control of Gametophytic Selection in Cocklebur, a 25-chromosome *Datura*, *Bot. Gaz.*, vol. 90, no. 4, pp. 366–383 (Dec.).

1931

43. Blakeslee, A. F, Extra Chromosomes, a Source of Variations in the Jimson Weed, *Smithsonian Report for 1930*, *Pub. 3096*, pp. 431–450 (13 plates).

1932

44. Bergner, A. D., and A. F. Blakeslee, Cytology of the *Ferox-quercifolia-stramonium* triangle in *Datura*, *Proc. Nat. Acad. Sci.*, vol. 18, no. 2, pp. 151–159 (Feb.).

45. Buchholz, John T., C. C. Doak, and A. F. Blakeslee, Control of Gametophytic Selection in *Datura* through Shortening and Splicing of Styles, *Bull. Torrey Bot. Club*, vol. 59, pp. 109–118 (Mar.).

46. ——— and A. F. Blakeslee, Pollen-Tube Growth in Primary and Secondary $2n + 1$ *Daturas*, *Amer. Jour. Bot.*, vol. 19, pp. 604–626 (July).

47. Blakeslee, A. F., The Species Problem in *Datura*, *Proc. 6th Internat. Cong. Genetics*, vol. 1, pp. 104–120.

1933

48. Bergner, A. D., S. Satina, and A. F. Blakeslee, Prime Types in *Datura*, *Proc. Nat. Acad. Sci.*, vol. 19, no. 1, pp. 103–115 (Jan.).

49. Blakeslee, A. F., A. D. Bergner, and A. G. Avery, Methods of Synthesizing Pure-Breeding Types with Predicted Characters in the Jimson Weed, *Datura stramonium*, *Proc. Nat. Acad. Sci.*, vol. 19, no. 1, pp. 115–122 (Jan.).

1934

50. Cartledge, J. L., and A. F. Blakeslee, Mutation Rate Increased by Aging Seeds as Shown by Pollen Abortion, *Proc. Nat. Acad. Sci.*, vol. 20, no. 2, pp. 103–110 (Feb.).

51. Blakeslee, A. F., and S. Satina, Les Plantes diffèrent-elles des animaux par des gamètes léthals? *Compt. Rend. Acad. Sci.*, vol. 198, no. 8, pp. 768–770 (Feb.).

52. Bergner, A. D., and A. F. Blakeslee, Cytology of a Translocation of the 1·2 Chromosome in *Datura*, *Bull. Torrey Bot. Club*, vol. 61, no. 4, pp. 197–209 (Apr.).
53. Blakeslee, A. F., New Jimson Weeds from Old Chromosomes, *Jour. Hered.*, vol. 25, no. 3, pp. 80–108 (Mar.).
54. Sinnott, Edmund W., Helen Houghtaling, and Albert F. Blakeslee, The Comparative Anatomy of Extra-Chromosomal Types in *Datura stramonium*, Carnegie Inst. Wash. Pub. No. 451, iii + 50 pp., 19 gravure pls. 28 text-figs.
55. Blakeslee, A. F., and A. G. Avery, Three Genes Located in the 21·22 Chromosome of the Jimson Weed, *Jour. Hered.*, vol. 25, pp. 393–404 (Oct.).
56. Bergner, A. D., J. L. Cartledge, and A. F. Blakeslee, Chromosome Behavior Due to a Gene Which Prevents Metaphase Pairing in *Datura*, *Cytologia*, vol. 6, pp. 19–37 (Dec.).

1935

57. Satina, S., and A. F. Blakeslee, Cytological Effects of a Gene in *Datura* Which Causes Dyad Formation in Sporogenesis, *Bot. Gaz.*, vol. 96, pp. 521–532 (Mar.).
58. Buchholz, J. T., A. F. Blakeslee, and A. G. Avery, Pollen-Tube Growth of a Translocation of the 1·2 Chromosome in *Datura*, *Proc. Nat. Acad. Sci.*, vol. 21, no. 4, pp. 208–213 (Apr.).
59. Cartledge, J. L., and A. F. Blakeslee, Mutation Rate from Old *Datura* Seeds, *Science*, vol. 81, no. 2107, pp. 492–493 (May).
60. Satina, S., and A. F. Blakeslee, Fertilization in the Incompatible Cross *Datura stramonium* × *D. metel*, *Bull. Torrey Bot. Club*, vol. 62, pp. 301–312 (June).
61. Bergner, A. D., and A. F. Blakeslee, Chromosome Ends in *Datura discolor*, *Proc. Nat. Acad. Sci.*, vol. 21, no. 6, pp. 369–374 (June).
62. Cartledge, J. L., M. J. Murray, and A. F. Blakeslee, Increased Mutation Rate from Aged *Datura* Pollen, *Proc. Nat. Acad. Sci.*, vol. 21, no. 11, pp. 597–600 (Nov.).
63. Buchholz, J. T., L. F. Williams, and A. F. Blakeslee, Pollen-Tube Growth of Ten Species of *Datura* in Interspecific Pollinations, *Proc. Nat. Acad. Sci.*, vol. 21, no. 12, pp. 651–656 (Dec.).

1936

64. Blakeslee, A. F., A. D. Bergner, and A. G. Avery, A New Method of Synthesizing Pure-Breeding Types with Extra Chromosomal Material in *Datura*, *Amer. Naturalist*, vol. 70, pp. 255–257 (May–June).
65. Cartledge, J. L., L. V. Barton, and A. F. Blakeslee, Heat and Moisture as Factors in the Increased Mutation Rate from *Datura* Seeds, *Proc. Amer. Philos. Soc.*, vol. 76, pp. 663–685 (Nov.).
66. Buchholz, J. T., and A. F. Blakeslee, Genes from Radium Treatment Affecting Pollen-Tube Growth in *Datura*, *Genetics*, vol. 21, pp. 731–751 (Nov.).

1937

67. Satina, S., A. F. Blakeslee, and A. G. Avery, Balanced and Unbalanced Haploids in *Datura*, *Jour. Hered.*, vol. 28, pp. 193–202 (June).
68. Blakeslee, A. F., A. D. Bergner, and A. G. Avery, Geographical Distribution of Chromosomal Prime Types in *Datura stramonium*, *Cytologia*, Fujii jubilee vol., pp. 1070–1093 (Aug.).
69. ——— Dédoublement du nombre de chromosomes chez les plantes par traitement chimique. *Compt. Rend. Acad. Sci.*, vol. 205, no. 11, pp. 476–479 (Sept.).
70. Satina, S., and A. F. Blakeslee, Chromosome Behavior in Triploids of *Datura stramonium*, I: The Male Gametophyte, *Amer. Jour. Bot.*, vol. 24, no. 8, pp. 518–527 (Oct.).
71. ——— ——— Chromosome Behavior in Triploid *Datura*, II: The Female Gametophyte, *Amer. Jour. Bot.*, vol. 24, no. 9, pp. 621–627 (Nov.).
72. Blakeslee, A. F., and A. G. Avery, Methods of Inducing Doubling of Chromosomes in Plants, *Jour. Hered.*, vol. 28, no. 12, pp. 393–411 (Dec.).

1938

73. ——— Colchicine, *Teaching Biologist*, vol. 7, no. 4, pp. 52, 61 (Jan.).
74. ——— Studies in the Behavior of Chromosomes, U. S. Dept. of Agric. Yearbook Separate No. 1605, pp. 1–35.
75. Satina, S., A. F. Blakeslee, and A. G. Avery, Chromosome Behavior in Triploid *Datura*, III: The Seed, *Amer. Jour. Bot.*, vol. 25, no. 8, pp. 595–602 (Oct.).

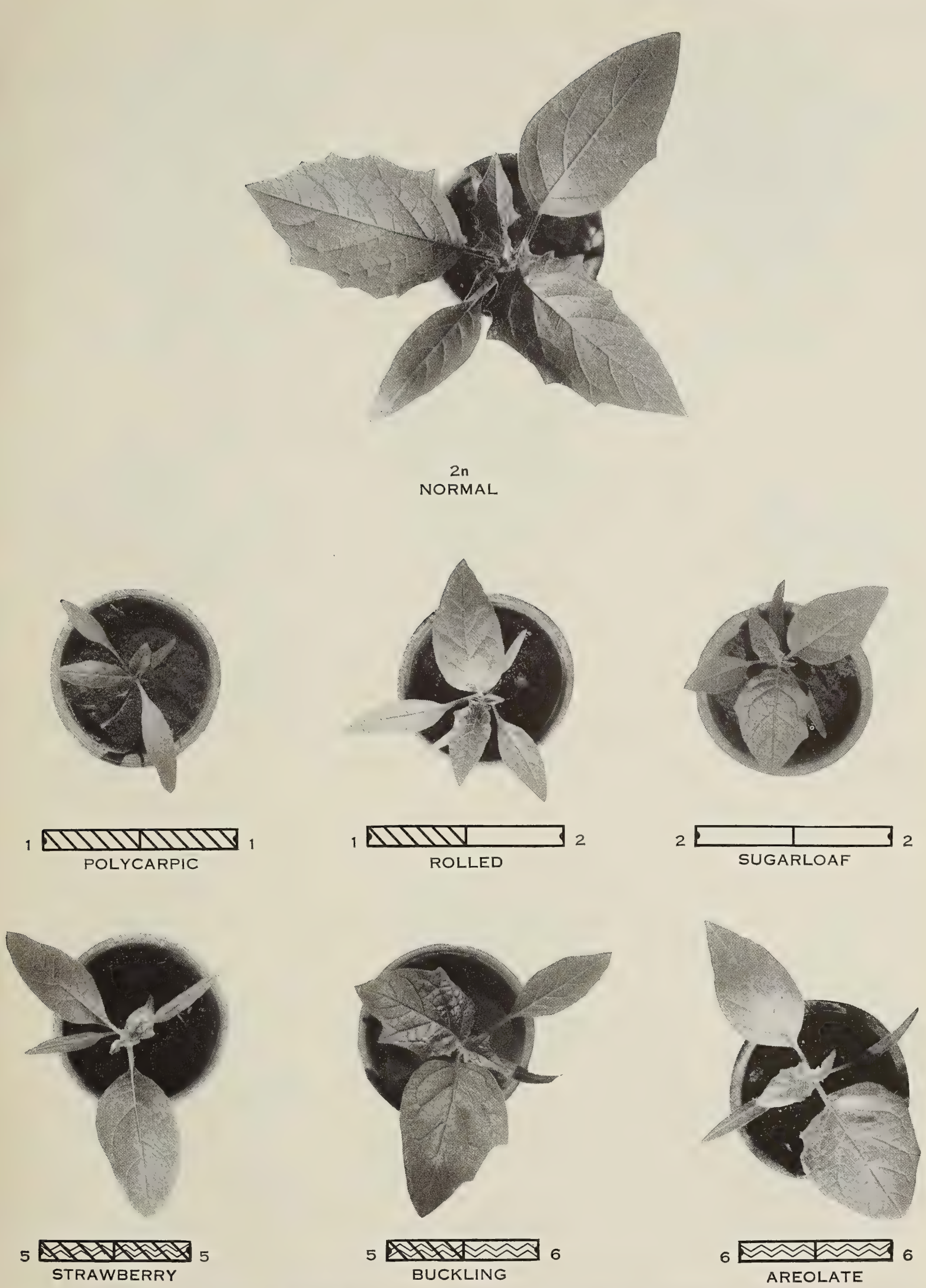

2n
NORMAL
1
1
POLYCARPIC
1
2
ROLLED
2
2
SUGARLOAF
5
5
STRAWBERRY
5
6
BUCKLING
6
6
AREOLATE

Plate 1. Above, normal $2n$ seedling in 3-inch pot. Below, seedlings of two primary $2n + 1$ types with two secondaries of each. Beneath each seedling is the common name and numbered model of the extra chromosome.

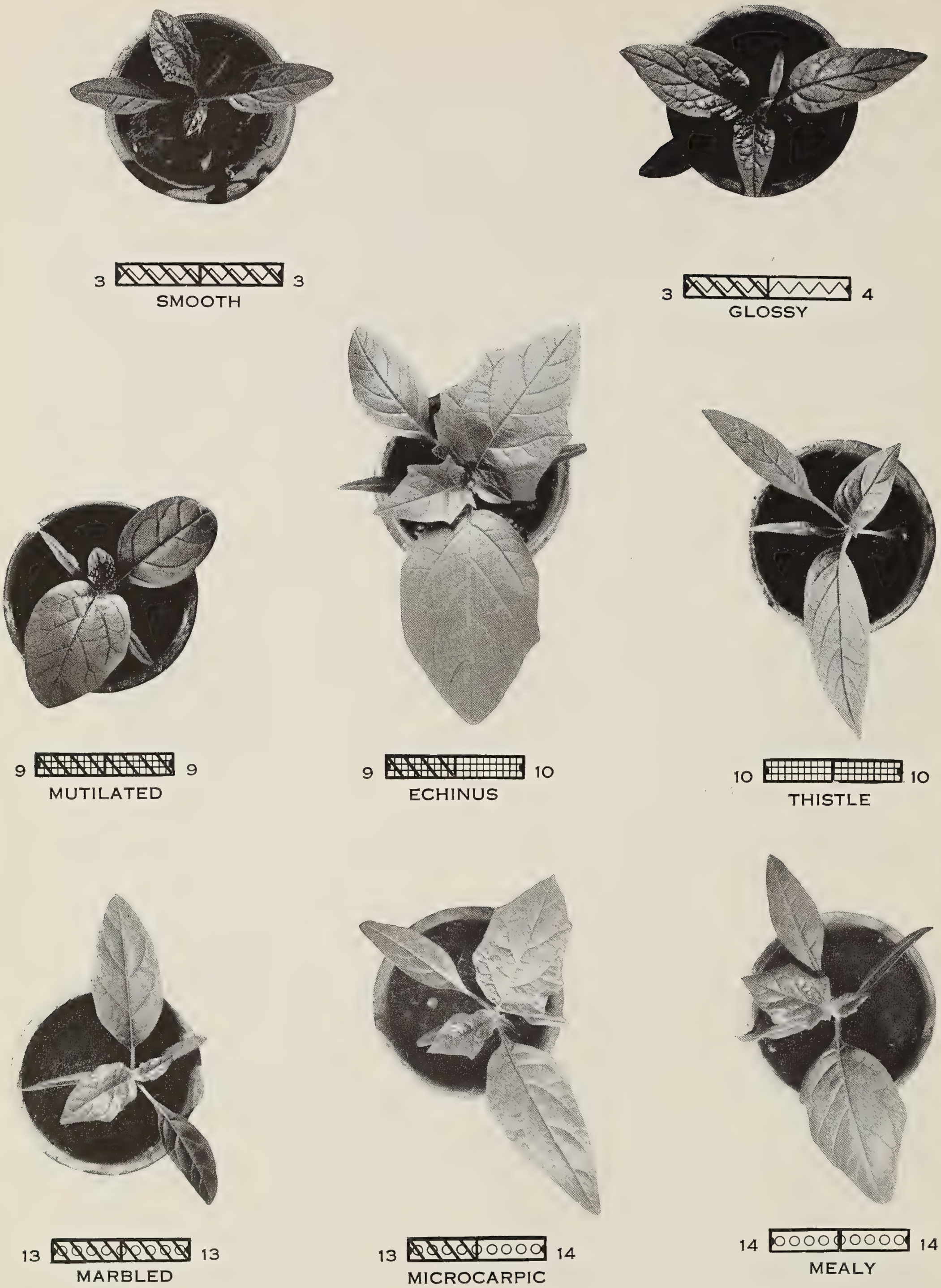
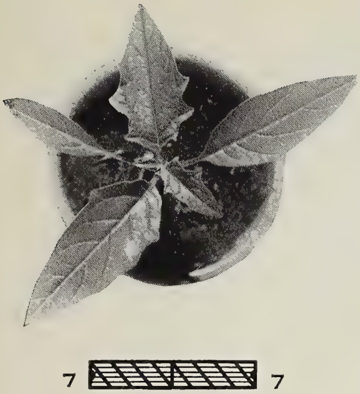
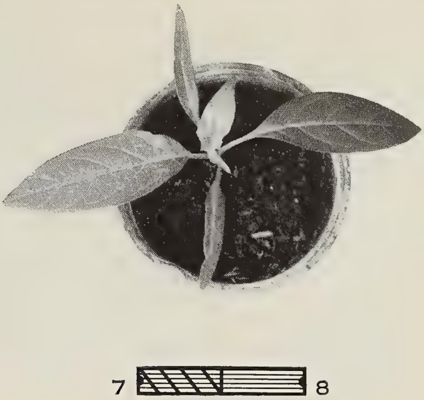

3 3
SMOOTH
3 4
GLOSSY
9 9
MUTILATED
9 10
ECHINUS
10 10
THISTLE
13 13
MARBLED
13 14
MICROCARPIC
14 14
MEALY

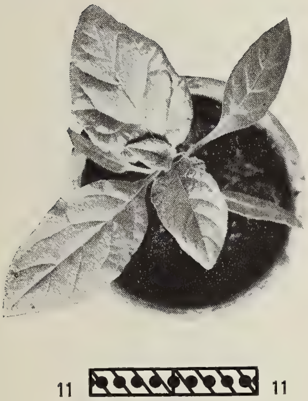
Plate 2. Above, seedling of a primary and its single secondary. Below, two primaries with two secondaries of each.



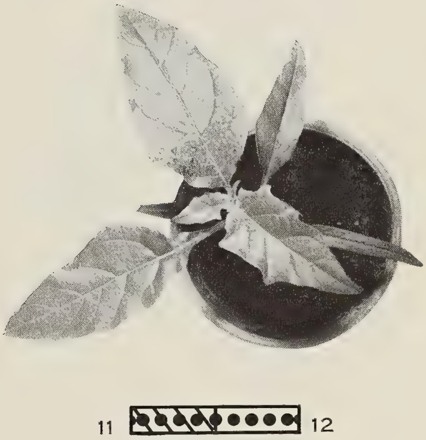
UNDULATE



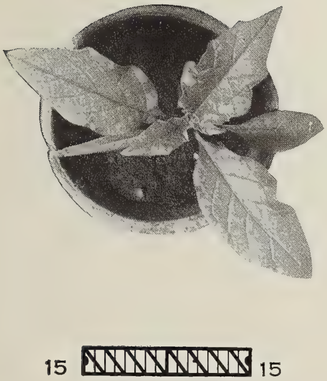
ELONGATE



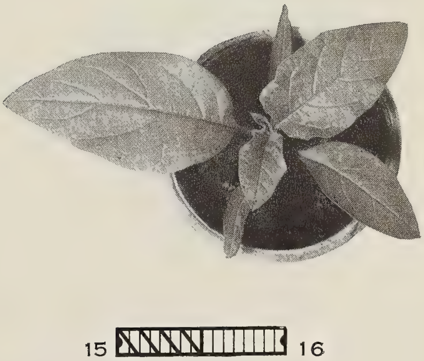
WEDGE



COCKLEBUR

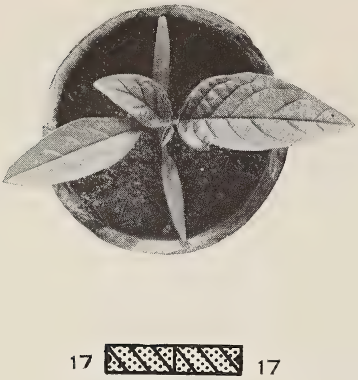


SCALLOPED

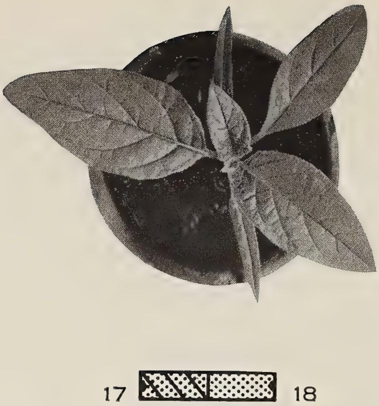


REDUCED

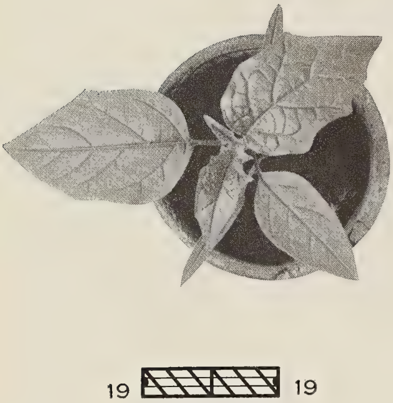
Plate 3. Seedlings of three primaries on right with their single secondaries on left.



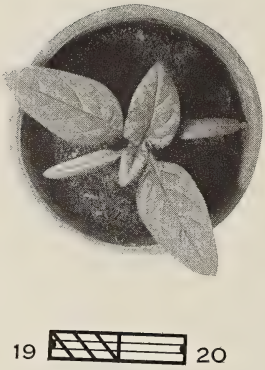
DWARF



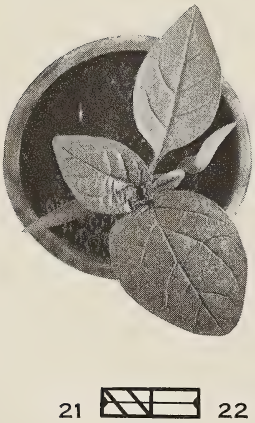
POINSETTIA



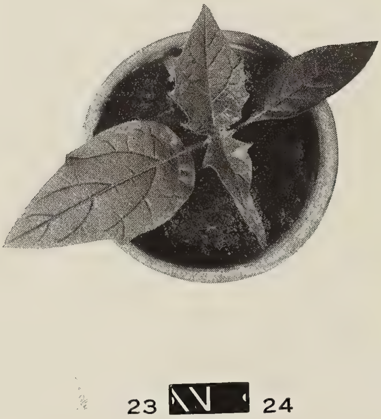
DIVERGENT



SPINACH

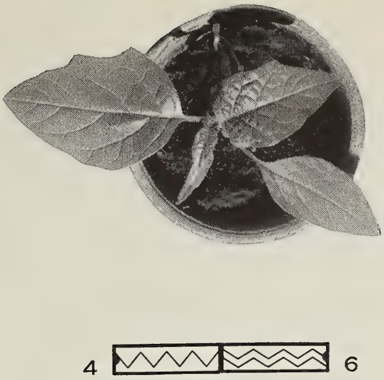


GLOBE

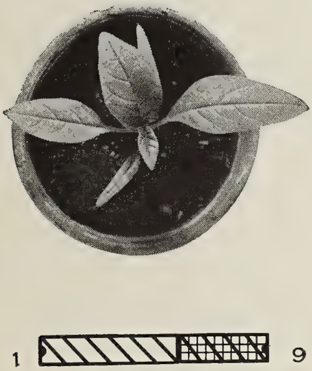


ILEX

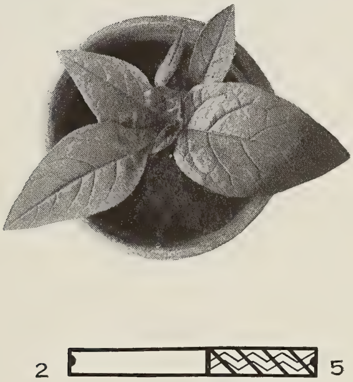
Plate 4. Above, seedlings of two primaries on right with their single secondaries on left. Below, seedlings of two primaries for which no secondaries are known.



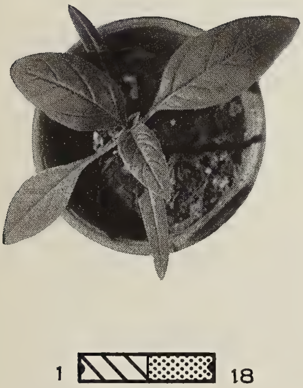
MAPLE



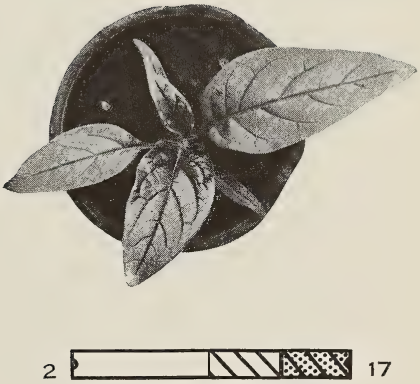
HEDGE



PINCHED



WIRY



DWARF-SUGARLOAF

Plate 5. Seedlings of tertiary $2n+1$ types. Above, single tertiary. Middle row, tertiaries which form the compensating type Nubbin. Bottom row, tertiaries from Prime Type 2.

WHAT AND WHY IS LEUKEMIA?—A MEDICO-BIOLOGICAL STUDY ON MICE

E. C. MACDOWELL

Department of Genetics
The Carnegie Institution of Washington

The biological interest in this project depends on the light it may throw on normal life processes; the medical interest, on the elimination of abnormal processes. In approaching these objectives simultaneously, the significance of observations is broadened.

The Leukemic Strain

The opportunity to undertake this cooperative project was provided by the antecedent development of a colony of numerous genetically uniform strains of pedigreed mice. The breeding in this colony was carried on in such a way that the original genetic variability was progressively eliminated. The system was the simple one of always breeding brother by sister. While inbreeding brings to light undesirable latent traits, once these are eliminated, inbreeding may be continued indefinitely without harm. This practice had been in effect for six years when the active study of leukemia was started in 1928. By that time the genetic homogeneity within each of the strains had reached a very high degree, as indicated by the pedigrees showing about eighteen generations of brother by sister matings. At the present time some strains are in the forty-eighth generation of this inbreeding. In one of these strains a surprising number of animals was found at autopsy to have enormous spleens, large livers, and swollen lymph nodes. They were dying with leukemia—in most cases, of the lymphatic type. On the basis of this strain the project was developed. Interpretable results were immediately obtained.

The statistical study of over 700 mice of this strain held until death gave the incidence of leukemia as 90 per cent. Starting with genetic uniformity, the difference between the 90 per cent that died with leukemia and the 10 per cent that died from some other cause could not be genetic; extrinsic influences were deciding into which group a given mouse would fall. In other words, the inherited tendency to leukemia was strong enough to become manifest under the conditions experienced by 90 per cent of the mice. All the mice were apparently subjected to the same conditions, yet among mice from the same litter, spending all their lives in the same box, one might develop leukemia at six months, others at different ages up to sixteen months, and one might die in old age without any signs of leukemia. The evidence of genetic uniformity was substantiated by the breeding records, which show that the incidence of leukemia was as high among the offspring of mice failing to develop leukemia as among the offspring of mice in which leukemia did develop. Given genetic uniformity, the rôle of non-genetic variables is brought to light.

Genetic uniformity of a strain means that all germ cells are genetically alike; all carry the same genes. If two different strains, each genetically uniform, are crossed, only one combination of genes is possible, so that all the mice in the first hybrid generation will be genetically alike. Thus when males from the leukemic strain were mated with females from a non-leukemic strain, all the young received the same combination of genes and variations in the results were due to extrinsic influences, as in the case of the pure-bred leukemic strain. However, in this first hybrid generation the incidence of spontaneous leukemia was reduced by one-half. With the genes of the leukemic strain contributed by one parent instead of by both, the chances of leukemia developing were halved. The range of variations in extrinsic influences remained the same, but the particular conditions necessary for leukemia were changed. These required conditions occurred with a frequency of 45 per cent, as compared with 90 per cent.

Reducing the strength of the genetic influences has increased the dependence upon extrinsic influences. This illustrates the constantly varying relation between heredity and environment, which provides endless material for the futile discussions between those who prefer to believe heredity supreme and those who are better satisfied by the idea that environment is all-important. Considering the strong human taboo against inbreeding, it seems probable that the rôle of extrinsic factors in the appearance of leukemia in man is far more comparable to that in the hybrid mice than to that in the pure-bred leukemic strain.

To maintain genetic uniformity requires the continuously successful conduct of a colony of pedigreed mice, based on long experience; to insure against confusion of identity of individual mice and against uncertainty of parentage requires special technique. Further, since non-genetic variables play a part, all animals should be raised under as nearly identical conditions as possible. For these reasons all the mice used in this entire project have been bred at the Department of Genetics at Cold Spring Harbor.

Leukemia, A Neoplastic Growth

Given this strain (C58) and others of different genetic constitution for comparison, the determination that leukemia is in fact a neoplastic growth was made simply and clearly. Leukemia differs from cancer in that the tissue involved gives rise to white blood cells, so that instead of the new-growth being primarily concentrated at the site of origin, the leukemic cells move freely through the body, multiplying as they go. The cells enlarge different parts, but their continuous multiplication is not dependent upon the formation of supporting structures by adjacent normal tissues; in this way leukemia differs from most tumors, which depend upon a non-malignant stroma carrying blood vessels. The name *leukemia* refers to the large numbers of these immature cells that usually appear in the peripheral blood stream, but other tissues, including the spleen, are usually heavily infiltrated before the cells appear in the blood.

The first stage of spontaneous leukemia. This general dissemination of leukemic cells has confused interpretation; many have thought that this indicated a disease of systemic origin. But the evidence now indicates that the very first stage of leukemia is a focal proliferation from reticulum cells. The sites have no fixity of position and are independent of normally organized centers of white blood cell formation. Leukemic cells are not blood cells that have become changed for some reason; they are a special race of cells having independent origin. From the centers of origin, migrations of immature but slowly and continuously differentiating cells invade any or all parts of the body. This migration obliterates all trace of the point of origin. Indeed the only possible means of determining this point was to watch the actual process at the time of origin. This was not hard to do in animals having a 90 per cent chance that leukemia would appear spontaneously. As a matter of fact, the ease of finding the earliest stages suggests that all starts do not necessarily progress to the point of developing the recognized signs of leukemia. That a marked degree of individuality has been found in frank cases of leukemia from this strain gives some support to the possibility that such differences at a very early stage may determine whether the wild growth of these cells is compatible with the condition of the animal at the time.

These early stages in themselves bear no evidence of any connection with leukemia. Their connection was first indicated by their appearing only in a series of 100 non-leukemic mice six months old or older from the leukemic strain. All organs were studied microscopically *in toto* by means of serial sections. The rate of cell dispersal, once the process is started, is so rapid that intermediate stages connecting these centers with early stages of recognizable leukemia were practically lacking. The identification of these unorganized proliferations of reticulum cells as the source of populations of leukemic cells depends on the integration of cytological and metabolic studies made on tissue removed at biopsy. With biopsy material removed at regular intervals from the same mouse the course of events in the same animal could be followed closely.

Metabolic rates of lymph nodes infiltrated with leukemic cells had previously been found to be high as compared with normal nodes. As expected, mice from several different pure-bred strains showed a decline in metabolic rates with increasing age; in this strain metabolic rates were found to rise with age. These observations were made on different mice at different ages since normal lymph nodes are very small and the respirometers generally used for studying metabolic rates require the use of all the nodes of a normal mouse. In order to determine the metabolic rates of a single normal lymph node, a special type of respirometer was developed. After testing the reliability of this new apparatus, and developing the necessary surgical technique, a series of 24 mice was carried through nine biopsies each, at intervals of a month; a tenth observation was made by sacrificing the animal. On each bit of tissue removed, both metabolic and cytological observations were made.

Since one node gives only a small sample of the conditions in the whole mouse, the irregularity of successive observations upon the same mouse was less surprising than the general consistency of the series from the same individual. Whenever the disorganized focal proliferations of reticulum cells were found, the metabolic rates were raised and usually at the next biopsy frank leukemia was found.

The above observations support the conclusion that leukemic cells arise independently of all normal lymphoid cells or organization. This has a bearing on the nature of the causative factors. If, as believed by some, an infection such as a germ or a virus were responsible, the first step would be the increased activity of organized lymph centers, and the disease would be systemic from the start instead of focal. Even after the leukemic population is frankly manifest, the relation to normal lymph nodes is invasion, not stimulation.

Transplantability of leukemic cells. Once the leukemic population is established it acts as an invasion of parasitic cells. The independent parasitic nature of populations of leukemic cells is emphasized by their ability to grow indefinitely as long as they are in a living, susceptible mouse. When a sample of a leukemic population is transplanted into a normal mouse genetically the same, the cells will continue to grow in the second mouse. Earlier failures to obtain this result in various laboratories probably resulted from the failure to use hosts genetically the same as the original spontaneous case. In this study, transplanting from a spontaneous case into hosts with various differing genetic constitutions has been repeatedly tried, in all cases unsuccessfully. Leukemic cells from more than forty spontaneous cases have been transplanted into young hosts of the same genetic constitution and in every case the leukemic cells have continued to grow in practically 100 per cent of the host mice. This process can be repeated indefinitely by removing a sample of leukemic cells from the transplanted mouse each time before it is killed by the cells. To be able to maintain populations of leukemic cells by this method of transplantation makes possible experiments on the nature of the change that places these cells outside the control which normally regulates the growth of such cells according to body needs.

Leukemic growth independent of bacteria. Transplantation experiments show that growth of these leukemic cells is not due to the presence of bacteria; test cultures were made at each transfer for a long series of transfers, and were always negative for bacteria. The examination of many thousands of sections of leukemic lesions has failed to disclose any bacteria. This, of course, does not say that bacteria may not be associated with leukemia, for they undoubtedly may be so associated, but it does say that leukemic growth proceeds in their absence. This has special significance in discounting press reports announcing cure of leukemia by antibacterial sera.

A leukemic agent not separable from living cells. Other experiments have attempted to extract some material from living leukemic cells that might induce leukemic activity in other cells. The cells have been killed in various ways—by heat, cold, grinding, hypo- and hypertonic solutions; various methods of filtration have been employed in air and under nitrogen; in every case in which living cells were not present leukemia failed to reappear. By extracting some substance from a colony of parasitic amœbæ it would be impossible to reproduce the colony in another host; apparently the populations of leukemic cells are similar to amœbæ in this respect.

Leukemic cells from leukemic cells. In one series of transfers starting from a spontaneous case, leukemic cells have been passed through 800 hosts in succession, over a period of ten years—several times as long as the life span of a mouse. And yet the evidence is strong that the cells of the present leukemic population are direct descendants, by an unbroken series of cell divisions, of the fertilized egg that developed into the mouse in which the leukemic differentiation from reticulum cells took place spontaneously.

Just as the specialized lymph organs do not contribute to the leukemic population in a spontaneous case, so the hosts into which leukemic populations are transplanted neither contribute cells to the leukemic population nor exhibit any increase in activity of lymphoid tissues as a result of the leukemic cells. If a virus or germ were responsible for the continuous growth of the leukemic population, one would expect an infection of the host's own cells, which would thereupon become leukemic. But exhaustive microscopic study establishes the purely invasive action of the transplanted cells, without activation of the host tissues. Different mice were sacrificed at regular intervals after transplantation of the leukemic cells; all critical organs were preserved and studied in their entirety by means of serial sections. Thus the history of the invading cells could be followed as the embryological history of germ cells is followed. In the terminal stage the organs and tissues might be so heavily infiltrated with invading cells that the original architecture was obliterated.

Leukemic cells depress metabolism of host tissues. Rather than stimulation, the effect of the leukemic cells upon normal tissue is actually depression. Before the transplanted leukemic cells demonstrably invade a tissue its metabolic rate is depressed below normal; after the cells arrive the high metabolic rate characteristic of the particular leukemic population is found.

The opportunity to make this significant observation depends on the fact that transplanted populations of leukemic cells taken from different spontaneous cases show individual characteristics. These differences may appear in length of the interval between a standard dose of cells and death of the host; in average cell size; in relative distribution of the cells in different organs. While most leukemic populations invade spleen, lymph nodes, and blood stream, others may show a practically normal blood picture, or the peripheral nodes and spleen may remain largely uninfiltrated. This last characteristic found in one leukemic population permitted the metabolic study of uninfiltrated lymph nodes from a mouse harboring progressively growing leukemic cells.

In short, leukemic cells are neoplastic because they arise from definite foci, continuously multiply as immature cells, invade and infiltrate normal tissues; they form an autonomous growth transplantable into susceptible hosts, and from host to host indefinitely, without breaking the continuous succession of cell divisions. Different populations of leukemic cells show great variety in specific histological and physiological characteristics that persist through many transfers from host to host, and yet these characteristics are subject to temporary modification or lasting change. The glycolytic rates of tissues infiltrated with leukemic cells is very high. In all these ways leukemia is paralleled by tumorous growths that are unquestionably accepted as neoplastic. And yet among pathologists the nature of leukemia is still considered debatable. In man there occur conditions that are easily confused with leukemia such as appears in strain C58, and yet do not behave as neoplasia. Given the infinitely greater variability in man of both genetic and extrinsic factors, it is entirely possible that intermediate conditions may occur in which an essentially neoplastic process is masked. The difficulty lies in the concept of a neoplasm as an entity depending for its expression only upon itself. The concept arising from these studies points to neoplastic growth depending on a balance between various factors; conversely, normal growth control depends upon a balance of different factors and various causes may upset this balance and yield uncontrolled growth. Throughout these studies a great similarity of malignant leukemic cells to normal lymphocytes, both morphological and physiological, has become more and more evident.

Resistance Induced in Susceptible Hosts

Resistance induced by leukemic cells. That a depressing effect precedes the arrival of the leukemic cells is an indication that the presence of the active leukemic population in some way reacts upon the host; some humoral response is concerned. Either a toxic substance has been given off by the leukemic cells, or the host is induced to form some material. In any case a reaction does take place between the cells and the host that defensively modifies the host. If the number of cells transplanted is very small, this defensive modification of the host may result in the destruction of all the leukemic cells and the survival of the host. To transplant a very small number of cells, such as 1/4,000,000 (theoretically 20 cells) of the standard dose, and have the mouse survive, might lead one to doubt whether the host had in fact received any leukemic cells. But this is not the case, for the mouse has been changed. If it is now given a larger dose of cells it will survive, and after a few more doses, each time with increasing numbers of cells, the resistance of the host will overcome massive doses of leukemic cells given repeatedly.

Why is the first stage of a spontaneous case not suppressed? With these facts in mind, the question may be asked, how does a case of spontaneous leukemia ever get started? Why do not the first leukemic cells call forth the defensive reaction that results in their destruction?

One suggestion is that the toxic effect of the spontaneous cells is too weak to call forth the defensive reaction of the animal. When transplanted for the first time the leukemic cells do not kill the hosts before 20 to 90 days; after a long series of transfers from mouse to mouse, the virulence may be so increased that the same number of cells will kill in 3 to 4 days. This increase in virulence undoubtedly means an increase in the toxic or depressing effect on the host; the hosts die before the growth of the leukemic population has caused any marked enlargement of organs, whereas in spontaneous cases leukemic growth usually causes enormous organ enlargement. Mechanical blockage is not the cause of death from highly virulent leukemic cells.

If a spontaneous case gets started because the first cells are too weak to call out the defensive reaction, it should be possible to prevent the appearance of a spontaneous case by first calling out the defensive reaction by highly virulent transplanted cells. A young mouse, immunized by increasing doses of highly virulent leukemic cells, should be able to resist the growth of leukemic cells that would otherwise appear spontaneously later in life. Over 100 young mice of the leukemic strain were immunized. To insure the persistence of this immunity, massive doses of highly virulent cells of a transplantation line were given monthly throughout their lives. At no time was there any suggestion of any weakening of the immunity. In spite of this, spontaneous leukemia, which is histologically distinguishable from the well-known transplanted cells, appeared as frequently as in unimmunized mice of this strain, namely in 90 per cent of the animals.

Spontaneous leukemic cells immunologically different from long-transferred cells. Paradoxically, the immunity of these mice resists highly virulent leukemic cells and yet does not resist the spontaneous cells of low virulence. Transplanted leukemic cells of high virulence and spontaneous leukemic cells differ not merely quantitatively in virulence, but also immunologically. Such an immunological difference largely eliminates the assumption of a virus or germ or any other independently organized infective agent responsible for the properties of leukemic cells. Immunity strong enough to inactivate the large amount of such an agent that would be present in a massive dose of highly virulent cells would be able to inactivate the relatively small amount of the same agent appearing in the first stage of a spontaneous case.

The general failure of various investigators to apply successfully to spontaneous tumors methods that induce resistance to transplanted tumors has given rise to the belief that the unique relationship between an animal and its own cells, even when they have become malignant, renders immunization impossible. In the present instance, this interpretation is ruled out by experiments showing that the same failure persists when the spontaneous leukemic cells are taken from their own mouse and inoculated into an immunized host; the immune host dies as soon as the non-immunized controls. The failure of this immunity is not due to a unique relationship between cells and hosts, but to immunological difference between spontaneous and long-transplanted leukemic cells.

Induction of resistance to spontaneous leukemic cells, a possibility. With the refutation of the belief that the unique relationship between a mouse and its own malignant cells renders immunization ineffective, the field is again opened for search for methods of inducing resistance to malignant cells produced in spontaneous cases. But in this search the tests must be made with malignant cells directly from spontaneous cases instead of from well-established transplanted neoplasia.

The change in malignant cells that takes place during the course of many transfers involves antigenic properties not originally present. This conclusion provides a consistent interpretation for many results reported for transplanted tumors. The continuum responsible for the persistence of leukemic growth through an infinite number of cell generations depends upon a mechanism, on the one hand, derivable from normal reticulum cells, on the other hand, capable of major antigenic modification as a result of transfer from mouse to mouse.

The demonstration of the immunological difference resulting from transplantation reduces the immediate therapeutic interest in methods of inducing resistance to transplanted leukemic cells, but the opportunity to trace, step by step, the origin of the immunological changes during transfer, and to analyze the various phenomena presented by controllable transplanted leukemic cells, may prove the most significant lead for the eventual control of spontaneous cases.

Nature of Induced Resistance

Passive transfer of resistance without antibodies. Immunization by small doses of highly virulent leukemic cells parallels the conventional technique of inducing resistance to countless deleterious entities. The mechanism is described in terms of toxic antigens inducing the production of defensive antibodies by the individual. However, in the present case no antibody can be found in the serum. While this is surprising, it is in accord with studies on the self-induced immunity of certain transplantable tumors; this induced resistance to certain transplantable tumors has never been transferred passively to other hosts. Although there are no protecting anti-leukemic bodies in the serum of an immunized mouse, a susceptible host can be immediately protected against a lethal dose of leukemic cells by treatment with minced tissue from an actively immunized animal. This must be considered passive immunity, since a period of time is required for the establishment of all forms of active immunity. Genetically and histologically identical tissue from non-immunized mice gives no protection. The protection is due to some material resulting from the active immunization, but this material is so intimately associated with living cells that it does not appear in serum; antibodies are said to pass freely throughout the animal's body. Any similarity to antibodies as known is further denied by the ease with which the protecting principle can be destroyed mechanically. The simple process of forcing the suspension of minced immune tissue out of a syringe held firmly against the bottom of the vessel removes the protecting effect; this process also tears apart all the cells. The destruction of the living cells liberates certain enzymes that might inactivate the protecting substance; but when enzyme action was inhibited by cold, there was still no protection. Is the protecting substance a large molecule directly disrupted by the same pressures that disrupt the cells, or is protection due to a modification of some normal component of the cells, destroyed by the disintegration of the cells? Discovery of the nature of this protecting substance would be a direct clue to the nature of the antigenic material in the leukemic cells, but the curious difficulty experienced in separating this substance and the ease of destroying it raise the question of the possible existence of an entirely unrecognized mechanism of resistance.

Resistance induced by normal tissue. Further suggestion of an unknown principle of resistance to neoplastic growth is given by protection established by treatment with normal tissue. In the case of transplanted leukemia, a single treatment with certain normal tissue will induce resistance to a lethal dose of leukemic cells. The effectiveness of the normal tissue does not depend upon its histological nature, since various organs, both embryonic and adult, may be used; but a certain genetic constitution is necessary—just what genetic constitution in turn depends upon the strain of susceptible hosts and the line of leukemic cells. Before making the test there is no means of knowing the effect of normal tissue of a given genetic constitution, except that tissue of the same genetic constitution as the host is ineffective.

A second unknown mechanism of resistance? Can it be that some normal tissue may have immunological similarity to highly virulent leukemic cells! When used in treatment, this normal tissue does not resemble leukemic cells in showing progressive growth; it quickly dies. On the other hand, tissue of the same genetic constitution as the host may show considerable growth; for example, embryo skin, although finely minced, will develop follicles and hair, and yet give no protection. Further, the resistance established by one single treatment with normal tissue differs from that induced by leukemic cells in that it cannot be passively transferred to another host, and, while regularly delaying the progress of leukemic invasion, does not always give lasting resistance. Besides delayed cases of leukemia that may develop in a month or two in contrast with death of controls in a week, curious subcutaneous tumors may appear with the histological characteristics of lymphosarcoma. Such tumors have never been encountered in any other transplantations of leukemic cells. On transplantation into normal hosts these tumors give rise to leukemia; not merely leukemia, but leukemia with all the characteristics peculiar to the line of leukemic cells previously inoculated.

These results indicate an incomplete form of resistance which may restrain the leukemic processes for weeks, so effectively that no histological evidence of any leukemia can be found, and yet finally permit a grossly modified expression of the leukemic continuum. If the tumor phase always disappeared in the first transfer into a normal host, it might be considered a physiological response to the altered environment, but in 8 of the 36 cases tested the tumor reappeared in one or more of the subsequent transfers through untreated hosts and thereby the question is raised of some change in the leukemic continuum itself. Since the active resistance induced by leukemic cells can be passively transferred and has never given delayed leukemia or subcutaneous tumors, it is easier to consider that normal tissue induces a different mechanism, than that it is immunologically similar to leukemic cells. If these two types of resistance are different, it may be that two new sorts of resistance instead of one are to be anticipated.

The resistance following the normal tissue treatment indeed appears to be some form of passive resistance; that is, due to material brought in by normal tissue itself rather than produced by the host. However, this material is not immediately present in available form, for two or three days must pass before the resistance is established. In twenty-four hours there is never any suggestion of resistance; in forty-eight hours some of the mice die with the controls while others survive indefinitely. In three days all are resistant. The particular normal tissue that has been most extensively studied comes from a strain of mice in which spontaneous leukemia is very rare. At one time this seemed suggestive of a natural anti-leukemia property dependent upon this particular genetic constitution. But such an assumption was ruled out when it was discovered that this same normal tissue when used in hosts of another strain actually facilitated the growth of the leukemic cells to which the mice of this other strain were susceptible.

Concluding Generalizations

Throughout these experiments, the results depend upon the relative properties of individuals and cells. A cell with leukemic properties in one animal, on transplantation into an animal with a different genetic constitution, may behave as a normal cell. This host is called non-susceptible, meaning that it does not provide the environment required by the given cell for the manifestation of leukemic properties. But there is great variety in the requirements of different leukemic cell populations; an environment that does not permit certain cells to show their leukemic properties may be perfectly suited for other populations to show leukemic growth. *Leukemia results from certain cells in certain environments.*

But the environment of the leukemic cells itself is a complex depending upon the relation of the particular genes and the particular extrinsic conditions encountered by the animal. The given genes operating under one set of conditions may result in a mouse susceptible to leukemia of line I; the same genes operating under the influence of certain treatments yield an environment in which leukemic cells of line I can not grow.

Intrinsic and *extrinsic factors* of *varying relative influence are concerned at every point and at every level.*

Not only does the interaction of genes and their environments decide the origin of spontaneous leukemic cells, but the subsequent interaction of the leukemic cells and their environment decides whether or not the leukemic cells survive; surviving, the performance of the leukemic population may be temporarily modified, or a persisting change in its properties may be induced by the surroundings.

Many of these results are in terms of transplanted leukemia; while not in detail referable to leukemia occurring spontaneously, the connection is very close and serves as a controlled basis of interpretation. The leukemia that appears naturally, without transplantation or any special treatment, in mice of strain C58, parallels certain types of leukemia in man, and the parallel is exceedingly faithful. The general conclusion that at present may be carried as far as man is that, while heredity may play a rôle, even in purebred mice environment has unquestionable influence, and in hybrids the dependence upon extrinsic influences is greatly increased; that, while neoplastic growth in the proper conditions appears ungovernable, it also is subject to extrinsic conditions. In other words, that failure to discover appropriate means of controlling tumor growth by changing its environment does not mean that this is impossible.

Knowledge of the interaction of tissues upon each other, of normal upon normal, of malignant upon normal, and normal upon malignant, may provide the clue to the clinical control of new growths, as well as to the basic processes of differentiation in individual development.

Acknowledgments

In the furtherance of the studies here reported in part, the following investigators have made larger or smaller active contribution:

M. Bovarnick, Department of Bio-Chemistry, College of Physicians and Surgeons, New York.

A. Claude, Rockefeller Institute for Medical Research, New York.

M. D. Findley, Department of Genetics, Cold Spring Harbor.

J. S. Lawrence, Department of Clinical Medicine, University of Rochester Medical School.

M. R. and W. H. Lewis, Department of Embryology, Baltimore.

C. J. Lynch, Rockefeller Institute for Medical Research, New York.

A. Papenheimer, Department of Pathology, College of Physicians and Surgeons, New York.

A. Pollock, Mount Sinai Hospital, New York.

J. S. Potter, Department of Genetics, Cold Spring Harbor.

M. N. Richter, Department of Pathology, College of Physicians and Surgeons, New York.

G. Schwartzman, Mount Sinai Hospital, New York.

E. E. Sproul, Department of Pathology, College of Physicians and Surgeons, New York.

M. J. Taylor, Department of Genetics, Cold Spring Harbor.

J. Victor, College of Physicians and Surgeons, New York.

E. N. Ward, Department of Genetics, Cold Spring Harbor.

M. P. Wintersteiner, Department of Pathology, College of Physicians and Surgeons, New York.

S. Wylens, Department of Pathology, College of Physicians and Surgeons, New York.

T. F. Zucker, Department of Organic Chemistry, College of Physicians and Surgeons, New York.

REFERENCES

The following references have been selected as an introduction to more detailed publications:

Reports of the Director of the Department of Genetics, Carnegie Institution of Washington Year Book, 1929 et seq.

MacDowell, E. C., and M. N. Richter, Genetic Aspects of Mouse Leukemia, *Biol. Zentralbl.*, vol. 52, pp. 266–279, 1932; *Arch. Pathol.*, vol. 20, pp. 709–724, 1935; *Amer. Jour. Cancer*, vol. 36, pp. 85–101, 1936; *Jour. Heredity*, vol. 28, pp. 130–138, 1937.

Potter, J. S., and M. N. Richter, Continuity of Cell Lineage in Transmission Lines of Lymphatic Leukemia, *Arch. Pathol.*, vol. 15, pp. 198–212, 1933.

Richter, M. N., and E. C. MacDowell, Experiments with Mammalian Leukemia (Review), *Physiol. Rev.*, vol. 15, pp. 509–524, 1935.

Victor, J., and J. S. Potter, Metabolic Studies, *Amer. Jour. Cancer*, vol. 22, pp. 561–571, 1934; *Brit. Jour. Exper. Pathol.*, vol. 16, pp. 253–265, 1935; *Amer. Jour. Cancer*, vol. 32, pp. 554–560, 1938.

STUDIES ON WHITE BLOOD CELLS

WARREN H. LEWIS AND MARGARET R. LEWIS

Department of Embryology
The Carnegie Institution of Washington

The white blood cells of animals and man are the subject of a vast literature extending over many years. Many investigators representing various branches of the biological sciences (embryology, zoology, cytology, histology, physiology, biochemistry, pathology, and medicine) have sought by various methods to elucidate their origin, their structure, their functions, and their varying numbers in health and diseases. Most of the older observations were made on fixed and stained smears, spreads, and tissues, and many of the conclusions arrived at have been and are still quite divergent in many important and fundamental respects. The development of the tissue culture and other methods for the study of living cells both *in vitro* and *in vivo* has lead to many new investigations in this field. During the past twenty years we have, with the cooperation of some fifteen investigators from other institutions, also applied these methods to the study of the living white blood cells and their relatives. More than fifty papers concerning them have emanated from this laboratory; the titles are listed below.

The significant rôle which blood cells play in health and in disease makes a complete understanding of their origin, cytology, behavior, and function a matter of great theoretical and practical importance.

It has been known for a long time that there are in the blood stream of normal human beings five easily recognized types of white blood cells among a total of about 7000 per cubic millimeter. The neutrophiles constitute about 69 per cent, the lymphocytes about 22, the monocytes about 5, the eosinophiles about 3.5, and the basophiles about 0.5 per cent. Studies of the ordinary Wright stained smears used by many investigators for many years have never settled the much disputed point as to the presence in the blood stream of transition forms between the lymphocytes and the monocytes.

Although the white blood cells in the blood stream of the various species of mammals, birds, reptiles, amphibia, and fishes vary considerably as to the character of the different types, the normal percentage of each type and the number per cubic millimeter are fairly constant for any species. It is thus evident that the number of each type of blood cell in the blood stream is under rigid control of the organism as a whole, just as are the various other organs of the body. This suggests that each type of blood cell has certain specific metabolic functions in the body but more particularly within the blood stream itself.

The buffy layer found when blood is centrifugalized consists almost entirely of white blood cells. Its volume is about 1.5 per cent of the blood. Since the total volume of blood is about 6000 cc, the white blood cells if gathered together into one mass would form an organ of considerable size, 90 cc or grams, an organ as large as the pancreas, twice as large as the thyroid, and 10 times as large as the two adrenals. Even the basophiles would form an organ as large as the hypophysis or the parathyroids. In spite of the total mass of these cells little is known of their metabolic rôle in the animal mechanism.

There are in addition to those in the blood stream great numbers of these cells in the lymph nodes, the thymus, the bone marrow, the spleen, and the tissue spaces throughout the body. Certain other cells such as macrophages, mast cells, and plasma cells which are related to the blood cells also occur in these tissues.

The white blood cells have relatively short lives; there is a continual outpouring of new ones into the blood stream and a continual removal of the injured and decrepit ones.

The outpouring of leucocytes into the blood stream and the increase and decrease in number of the different types as well as changes in their character in various diseases have led to a great series of studies by many investigators concerning their origin from the hematopoietic organs such as bone marrow, lymph nodes, and spleen. Some authors maintain that all the white blood cells as well as the red blood corpuscles and the relatives of the former found in the tissues come either directly or indirectly from a common stem cell, the hemocytoblast. Others maintain that the stem cell of the granulocytes is different from that of the lymphocyte and monocyte. Some hold that the last two types are independent self-perpetuating strains of cells. Many variations of these theories have been promulgated.

The usual procedure for determining the origin and relationships of the blood cells by arranging a selected number of fixed and stained dead ones in a series is open to serious objection. The ideal method would be actually to follow the changes which living cells undergo. Tissue cultures seem to present such an opportunity. Difficulties were at once encountered in identifying the living cells. They are very different from dead and stained ones of the same kind. However, we have been able to build up criteria for the identification of some of them.

As a connecting link between the fixed and stained dead cells of tissues and the more or less free moving cells of cultures we have examined living spreads of various tissues with and without the use of vital dyes. In such preparations, where small pieces of normal, experimental, and pathologic tissues are pressed out into a thin layer between the cover glass and slide, one can identify many cells and get some idea of their characteristics and distribution.

Neutrophiles

Cultures of the buffy coat of the blood give a prompt and abundant migration of neutrophiles followed a few hours later by monocytes and small macrophages. The neutrophiles have a rapid and characteristic mode of migration but live as a rule only a few days.

It is in cultures of bone marrow that one would expect to obtain direct evidence of the change of stem cells into neutrophiles. Such cultures show migration of neutrophiles and of myelocytes and myeloblasts, which most investigators consider as the immediate forerunners of the neutrophiles. From a study of the motion pictures of such preparations it has been possible to demonstrate that the myeloblasts have a characteristic mode of locomotion quite different from that of the other cells. The myeloblasts and myelocytes undergo mitotic division in the cultures, giving rise to cells of the same type, not to more differentiated cells. Other types of cells present in the marrow, such as lymphocytes, monocytes, small macrophages, endothelial cells, and fibroblasts, also migrate out of the explants.

Unfortunately cultures of the type used at the present time have not proved to be suitable for the transformation of primitive stem cells into the adult forms, and so we have not been able to follow transformation of the myeloblasts into neutrophiles.

The Lymphocytes

Cultures of lymph nodes give prompt and abundant migration of lymphocytes. These cells have a rapid and characteristic mode of locomotion. As a rule they live only a few days and show no signs of division or change into other types of cells. Lymphocytes are also found in cultures of other organs and tissues, especially certain tumors, and under certain conditions they may live and move about for many days.

Cultures of lymph nodes from young animals show in addition to the small lymphocytes large lymphoid cells, the lymphoblasts, which histologists have identified as the cells that give rise to the lymphocytes. The lymphoblasts undergo mitotic division and move about with a characteristic speed and mode of locomotion, but their further transformation has not been observed in our cultures.

The initial outgrowth of lymphocytes is soon followed by the appearance of monocytes and small macrophages which thrive for many days, multiply, phagocytose, pinocytose, increase in size, and often fuse to form large multinucleated cells. Later, endothelial cells and connective tissue cells migrate out into the medium, multiply, and under suitable conditions can be cultivated indefinitely. Among these cells we have found no evidence of a reticular stem cell which gives rise to lymphocytes or lymphoblasts.

In vivo the lymphoblasts and lymphocytes multiply in the lymph nodes, and the myeloblasts and myelocytes in the bone marrow, but in tissue cultures of the blood-forming organs as well as of the buffy coat the surviving white blood cell is the monocyte. Until a suitable medium is found for the long-continued cultivation *in vitro* of the other types of white blood cells many problems concerning them will remain unsolved.

Monocytes, Macrophages, Epithelioid Cells

Our studies on the cultivation of the cells from the blood stream with and without the use of vital dyes have led to the observation of the change of living monocytes into macrophages and epithelioid cells and various transitions between them. The macrophages that arise in these cultures are identical with those in cultures of fixed tissues, and the epithelioid cells with those obtained in cultures of tubercular nodules. Repeated observations on cultures of blood of different species of animals from fish to man have confirmed this finding and established on a secure basis the transformation of monocytes into macrophages and epithelioid cells as well as the fact that these cells are fundamentally identical, the difference in appearance being due principally to the size of the ingested materials and degree of their ingestion. Actual observations on similar cultures showed that macrophages and epithelioid cells often fuse to form giant cells like those in tubercles. Observations on the living Kupffer cells of the liver, which were supposed by many to be of endothelial origin, and on the epithelial phagocytes of the lung, which were by some observers considered to be of epithelial origin, revealed that they are macrophage-epithelioid cells and thus of monocyte origin. Up to the time of these observations much uncertainty existed concerning the origin of the macrophages and epithelioid cells and various theories were advocated, based for the most part on the study of fixed and stained cells *in situ* in the tissues.

Monocytes and macrophages have been identified in all types of fixed tissue except the retina of the eye and the optic lobes of the brain. Their presence has been demonstrated in embryonic as well as adult tissue.

Concerning the origin of monocytes there are many opinions. The idea that they come from lymphocytes is maintained by several investigators, but repeated observations on living lymphocytes in cultures have failed to show an actual transformation of a lymphocyte under observation into a monocyte. The smallest monocytes are about the size of lymphocytes, but their mode of locomotion, pseudodia, and number of neutral red granules enable us to distinguish one type of cell from the other.

Monocytes and small macrophages are abundant in cultures of tissues. Under proper conditions they are active, multiply, live for long periods, and seem to be a self-perpetuating strain of cells which can probably be carried on indefinitely in cultures. They often exhibit phagocytosis and become large and sluggish when they have ingested large amounts of cell débris and dead cells. They also drink in large amounts of fluid (pinocytosis). Under certain conditions they assume forms somewhat like fibroblasts, but the claim made by some authors that they change into true fibroblasts seems to us very doubtful.

Some authors claim that monocytes and small macrophages arise from the endothelium, especially embryonic. Endothelium is common in cultures of both embryonic and adult tissues and may grow out extensively. We have, however, never seen any indication of the origin of monocytes or small macrophages from it.

Locomotion of Blood Cells

Locomotion of blood cells and their relatives in tissue cultures has attracted our attention for many years. The application of motion pictures to these studies has been of invaluable assistance. They have revealed processes that might not have been otherwise noted and differences of movements which are not appreciated when one follows living cells with the eye under the microscope. For example, the process we have called pinocytosis or drinking by cells was first seen in motion pictures of macrophages. This process is probably more important than phagocytosis. It had been going on for years under our eyes, but not until the drinking was speeded up by the motion pictures was it recognized. By comparing the movements of cells seen by the eye with the same or similar ones in motion pictures where the movements are speeded up from 24 to 32 or more times normal, it has been found that such cells as lymphocytes, lymphoblasts, neutrophiles, myeloblasts, monocytes, and macrophages have each a characteristic mode of locomotion.

The motion picture films have been utilized in quite a different manner for the study of some of these cells. By projecting one frame at a time and noting the change in position and in form of individual cells in each successive picture taken at stated intervals it was possible to determine the rate of locomotion of many cells photographed at the same time. It was surprising to find that lymphocytes migrate as rapidly as the supposedly more lively neutrophiles.

This method yielded a fact of even greater significance and one totally unexpected, namely, that the constriction ring of migrating cells does not move forward as the cell moves forward (see figure 1), but remains fixed as regards points outside the cell. It was now possible to compare the locomotion of these cells with that of the amœba, where many of the factors concerned were well known, and to determine that the same factors were at play in these cells.

The next step also came from an unexpected quarter, namely, from an attempt to explain the peculiar contortions seen in motion pictures of the dividing cells of a mouse sarcoma. It was found that the contortions could be most clearly explained on the assumption that such cells had a superficial layer which was in a continual state of contraction and that local variations in the strength of the contraction produced the constrictions and lobulations. It is well known to chemists that when colloids gel they automatically exert considerable contractile tension. Cytoplasm is a colloid, hence any gelation will presumably produce contractile tension, and local variations of the viscosity of the gel will produce corresponding

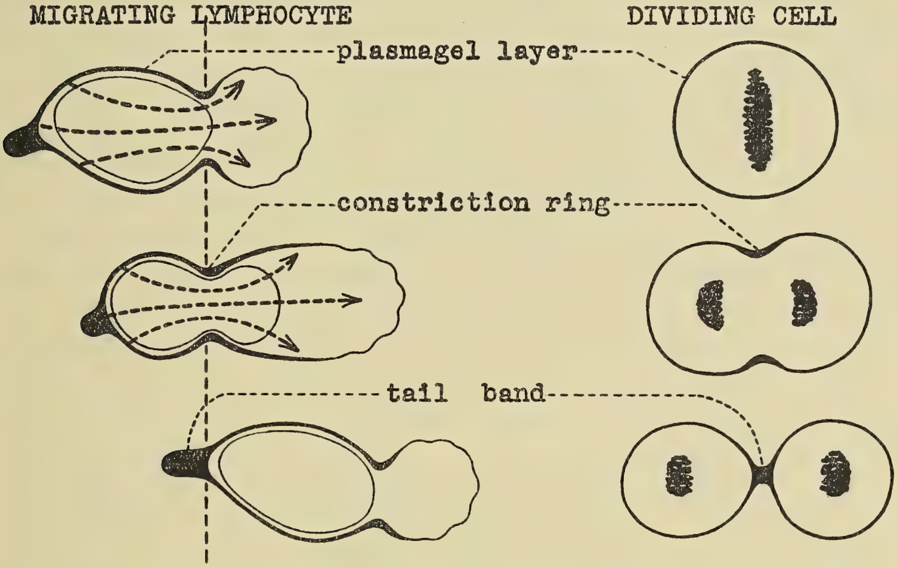


Fig. 1. A superficial plasmagel layer exerts continuous contractile tension, thrusts out a pseudopod at a thin area, and forces forward endoplasm, nucleus, and plasmagel as it goes into sol state. Continuous gelation at the base of the advancing pseudopod forms a new advancing plasmagel layer. A thick plasmagel band produces constriction ring because of increased contractile tension. It produces the gelated tail of the lymphocyte and pinches the dividing cell in two except for a gelated band which may persist for a short time.

local variations in the tension. The assumption that cells possess a superficial plasmagel layer which is automatically contractile supplies a missing link in the explanation of their locomotion and changes in form.

The diagram illustrates our conception of how the contraction of such a layer continually thrusts the cytoplasm and nucleus forward through a softened spot around which a constriction ring is developed by an increased contraction of a thickened band of this plasmagel layer. The constriction of the nucleus indicates that it exerts considerable tension. This ring serves as an important landmark since it does not move forward. The cytoplasm that is thrust forward through the ring undergoes at the surface a progressive gelation from the ring forward to form a new plasmagel layer. As the plasmagel layer posterior to the ring contracts and thrusts the nucleus and softer endoplasm forward, it becomes smaller and smaller as it gradually goes into solation or becomes less viscous. Finally the nucleus and all the endoplasm and plasmagel layer pass through the ring except a somewhat reduced tail and constriction ring. The constriction ring or what is left of it then contracts down and increases the length of the plasmagel tail. As the cell moves forward new rings form and the same process is repeated over and over.

The lymphocyte is about the simplest of all the cells we have so far examined, but the same general factors seem to apply to the locomotion of various other types of cells.

The accompanying figure also illustrates our conception of the action of a local thickened circular band of the superficial plasmagel layer around the equator of a bilateral dividing cell. Such a band would automatically exert greater contractile tension than the rest of the plasmagel layer and produce a constriction ring, similar to the one on the lymphocyte, which divides the cell in two. A small band usually connects the two daughter cells for a short time after division. It corresponds to the gelated tail of the lymphocyte. Thus some of the same factors which seem to be responsible for locomotion probably play a part in cell division.

Abnormal White Blood Cells

Since different types of white blood cells can be identified by peculiarities in their mode of locomotion especially when they are observed in motion pictures of tissue cultures, it seemed probable that tissue cultures and motion pictures of the various cells found in leukemias and other blood diseases might help to differentiate more clearly the types of cells involved. In collaboration with Dr. Rich of the Department of Pathology and Dr. Wintrobe of the Department of Medicine of the Johns Hopkins University, such studies are now in progress. Some of the leukemias have fairly typical cells and are easily diagnosed by the usual blood method. These show the same types of cells in the motion pictures.

Other types of this blood disease are characterizied by the presence in the blood stream of young forms normally present only in the blood-forming organs. These young forms are known as blast cells (lymphoblasts, myeloblasts, and possibly monoblasts). In stained blood preparations it is difficult to distinguish the young lymphoid, myeloid, and monocytic cells. In the motion picture studies, on the other hand, each of these cells has been found to exhibit a characteristic speed and mode of locomotion so that in many instances it has been possible to identify the type of blast cell present in the particular blood in question.

MALIGNANT WHITE BLOOD CELLS

Earlier work on tumors led to the conclusion that tumor cells are permanently altered cells which are able to multiply without control of the host. There is reason to surmise that the blood cells, under the proper stimulant, might take on malignant characteristics and that we may therefore, expect to find tumors composed of the primitive forms as well as the more adult stages of each kind of blood cell.

We have had 6 white blood cell tumors, 4 lymphosarcomata of lymph nodes, and 2 monocytomata of the spleen among 294 dibenzanthracene tumors induced in pure strains of mice. Only two of the lymphosarcomata were studied by means of cultures and motion pictures. The malignant cells of one tumor were lymphoblastic in character and those of the other were like large lymphocytes. The lymphoid cells migrated from the explants more slowly than normal ones and exhibited many mitoses. These cells showed the characteristics of malignant cells in that they were larger and had larger and more granular nuclei than normal ones. Each tumor has been transplanted through many generations and like the spindle-cell sarcomas induced by dibenzanthracene they are 100 per cent transplantable in mice of the strain of origin and not transplantable in other strains. The blood is not leukemic yet it contains malignant cells, for the injection of 0.8 cc of blood into a mouse will produce a characteristic lymphosarcoma.

The monocytomata arose in the spleen. One has only recently been found; the other has been transplanted subcutaneously through many generations and grows in 100 per cent of the mice of the strain of origin and not in mice of other strains. The growth of this tumor was not accompanied by characteristic changes in the blood of the host other than a slight increase in number of monocytes and the presence in the stained blood spreads of an occasional larger monocyte with slightly bluer cytoplasm. The injection of 0.5 to 0.8 cc of this blood subcutaneously into normal young mice of the susceptible strain, however, resulted in the growth of a monocytoma at the injection site.

The cells of this tumor were monocytic cells of the large epithelioid cell type. Motion pictures showed that they behaved like the large epithelioid cells seen in the tissue cultures of tubercle. They were phagocytic and ingested red blood cells and formed bilirubin crystals. Their nuclei were somewhat larger and had more chromatin granules and more nucleolar material than those of normal epithelioid cells.

Blood Picture of Tumor-bearing Mice

Not only may the different types of white blood cells themselves become malignant, but the growth of certain of the neoplasms of fixed tissue bring about changes in the blood of their hosts. Observations made of the blood and blood-forming tissues of mice bearing induced dibenzanthracene spindle-celled sarcomata and others with spontaneous mammary gland carcinomata revealed that the effect upon the blood-producing organs of the host involved the myeloid rather than the lymphoid or monocytic cells.

Some tumor-bearing mice exhibited only a slight neutrophilia in their peripheral blood while others exhibited a profound leucocyotisis with a marked increase in the number of polymorphonuclear cells. When these tumors were transplanted into normal young mice their growth was accompanied by an increase in the size of the spleen and a greater hyperplasia of myeloid tissue even than that exhibited by the host of the primary tumor.

The tumors that brought about the most profound myeloid hyperplasis in their hosts proved to be somewhat less strain specific than the majority of the tumors studied. Some of them grew to a large size in mice of alien strains although they later regressed, and a few of them following repeated implantation into mice of alien strains became so changed that they were thereafter transplantable into many strains of mice.

The cells present in the neutrophilic blood of the tumor-bearing hosts were not malignant, and the injection of this blood did not result in the development of tumors.

Mice bearing a monocytoma or a lymphosarcoma implanted with one of these myeloid-stimulating sarcomata developed a profound neutrophilia with large numbers of myeloid cells in the spleen and infiltrations of myeloid cells in the liver. The great hyperplasia of myeloid cells, however, did not prevent the presence in the blood stream of some of the specific malignant cells, and the injection of this myeloid blood resulted in the growth of a lymphosarcoma or a monocytoma according to the type of neoplasm present in the host. The presence of malignant lymphoid or monocytic blood cells does not prevent the formation of an overwhelming number of normal neutrophiles.

STUDIES ON WHITE BLOOD CELLS

Lewis, W. H., and L. T. Webster, 1921*a*. Migration of Lymphocytes in Plasma Cultures of Human Lymph Nodes, *Jour. Exper. Med.*, vol. 33, pp. 261–269.

Lewis, W. H., and L. T. Webster, 1921*b*. Giant Cells in Cultures from Human Lymph Nodes, *Jour. Exper. Med.*, vol. 33, pp. 349–360.

Lewis, W. H., and L. T. Webster, 1921*c*. Wandering Cells, Endothelial Cells and Fibroblasts in Cultures from Human Lymph Nodes, *Jour. Exper. Med.*, vol. 34, pp. 397–405.

Lewis, W. H., and C. C. McCoy, 1922. The Survival of the Cells after the Death of the Organism, *Johns Hopkins Hosp. Bull.*, vol. 33, pp. 284–293.

Beck, C. S., 1922. The Relative Distribution of Clasmatocytes in the Various Organs of the Seven-Day Chick Embryo, *Anat. Rec.*, vol. 24, pp. 61–66.

Rioch, D., 1923. The Morphology and Behavior of the Migratory Cells in Tissue Cultures of the Chick Spleen, *Anat. Rec.*, vol. 25, pp. 41–57.

Lewis, W. H., and G. O. Gey, 1923. Clasmatocytes and Tumor Cells in Cultures of Mouse Sarcoma, *Johns Hopkins Hosp. Bull.*, vol. 34, pp. 369–371.

Lewis, W. H., 1923. Hofbauer Cells (Clasmatocytes) of the Human Chorionic Villus, *Johns Hopkins Hosp. Bull.*, vol. 35, p. 183.

Rich, A. R., 1924. The Formation of Bile Pigment from Hæmoglobin in Tissue Cultures, *Johns Hopkins Hosp. Bull.*, vol. 35, p. 415.

Lewis, M. R., 1925*a*. The Formation of Macrophages, Epithelioid Cells and Giant Cells from Leucocytes in Incubated Blood, *Amer. Jour. Pathol.*, vol. 1, pp. 91–100.

Lewis, M. R., H. S. Willis, and W. H. Lewis, 1925. The Epithelioid Cells of Tuberculous Lesions, *Johns Hopkins Hosp. Bull.*, vol. 36, pp. 175–184.

Lewis, M. R., H. S. Willis, and W. H. Lewis, 1925. The Epithelioid Cells of Tuberculous Lesions, *Tubercle*, April 1925, pp. 1–7.

Lewis, M. R., 1925*b*. Origin of the Phagocytic Cells of the Lung of the Frog, *Johns Hopkins Hosp. Bull.*, vol. 36, pp. 361–375.

Lewis, M. R., 1925*c*. The Resemblance between the Kupffer Cells and the Mononuclear Leucocyte of the Frog Studied in vitro, (abstract) *Anat. Rec.*, vol. 29, p. 390.

Lewis, M. R., 1925*d*. Origin of the Phagocytic Cells of the Lung of the Frog, (abstract) *Anat. Rec.*, vol. 29, p. 390.

Lewis, M. R., and W. H. Lewis, 1925. Monocytes, Macrophages, Epithelioid Cells and Giant Cells, (abstract) *Anat. Rec.*, vol. 29, p. 391.

Lewis, W. H., 1925*a*. Cell Inclusions, Vital Dyes and the So-called "Segregation Apparatus," (abstract) *Anat. Rec.*, vol. 29, p. 391.

Lewis, W. H., and M. R. Lewis, 1925. The Transformation of Leucocytes into Macrophages, Epithelioid Cells and Giant Cells in Cultures of Pure Blood, (abstract) *Amer. Jour. Physiol.*, vol. 72, p. 196.

Lewis, W. H., 1925*b*. Engulfment of Living Blood Cells by Others of the Same Type, *Anat. Rec.*, vol. 31, pp. 43–49.

Lewis, M. R., 1925*e*. A Study of the Mononuclears of the Frog's Blood, *Arch. f. exper. Zellforsch.*, vol. 2, pp. 228–243.

Lewis, W. H., 1925–1926. Transformation of Mononuclear Blood Cells into Macrophages, Epithelioid Cells and Giant Cells, *Harvey Lectures*, vol. 21, p. 77.

Lewis, M. R., and W. H. Lewis, 1926. Transformation of Mononuclear Blood Cells into Macrophages, Epithelioid Cells and Giant Cells in Hanging Drop Cultures from Lower Vertebrates. Contr. to Embryol. No. 96, Carnegie Inst. Wash. Pub. No. 363, pp. 95–120.

Eliot, C. P., 1926. The Origin of the Phagocytic Cells in the Rabbit, *Johns Hopkins Hosp. Bull.*, vol. 39, pp. 149–157.

Lewis, W. H., 1926. Macrophages in Sterile Inflammation of the Deep Fascia of the Rat, (abstract) *Anat. Rec.*, vol. 32, p. 215.

Lewis, M. R., and H. B. Andervont, 1926*a*. The Malignant Cell of the Chicken Sarcoma, (abstract) *Anat. Rec.*, vol. 32, p. 237.

Lewis, M. R., and H. B. Andervont, 1926*b*. The Serial Transmission of Chicken Tumors by Means of Injections of the White Blood Cells and the Plasma, *Amer. Jour. Hygiene*, vol. 6, pp. 498–505.

Lewis, W. H., 1927*a*. Binucleate and Giant Cells in Tissue Cultures and the Similarity of the Latter to the Giant Cells of Tuberculous Lesions, *Tubercle*, April.

Lewis, W. H., 1927*b*. The Formation of Giant Cells in Tissue Cultures and Their Similarity to Those in Tuberculous Lesions, *Amer. Rev. Tuberculosis*, vol. 15, pp. 616–628.

Lewis, W. H., 1927*c*. Migration of Neutrophilic Leucocytes, *Arch. f. exper. Zellforsch.*, vol. 4, p. 442.

Lewis, W. H., 1927*d*. Sarcoma Cells, *Arch. f. exper. Zellforsch.*, vol. 5, pp. 143–156.

Lewis, W. H., 1928*a*. The Transformation of Mononuclear Blood Cells into Macrophages, Epithelioid Cells and Giant Cells, *Arch. f. exper. Zellforsch.*, vol. 6, pp. 253–259.

Rich, A. R., and M. R. Lewis, 1928. Mechanism of Allergy in Tuberculosis, *Proc. Soc. Exper. Biol. and Med.*, vol. 25, pp. 596–598.

Lewis, W. H., 1928*b*. Comparative Study of Fish Blood, *Carnegie Inst. Wash. Year Book No. 27*, pp. 279–280.

Lewis, W. H., 1929. Macrophages and Other Cells of the Deep Fascia of the Thigh of the Rat. Contr. to Embryol. No. 116, Carnegie Inst. Wash. Pub. No. 394, pp. 193–212.

Lewis, W. H., 1930. Moving Pictures of Normal Cells, (abstract) *Anat. Rec.*, vol. 45, p. 229.

Lewis, W. H., 1931*a*. On the Locomotion of Lymphocytes, (abstract) *Anat. Rec.*, vol. 48, p. 52.

Lewis, W. H., 1931*b*. Locomotion of Lymphocytes, *Johns Hopkins Hosp. Bull.*, vol. 49, pp. 29–36.

Lewis, W. H., 1931*c*. Pinocytosis, *Johns Hopkins Hosp. Bull.*, vol. 49, pp. 17–27.

Lewis, W. H., and M. I. Rubin, 1932. The White Cells in Cultures of the Buffy Coat of Human Blood, *Mt. Desert Isl. Biol. Lab. Ann. Rept., 34th season*, p. 32.

Lewis, W. H., 1932. The Locomotion Rate of Rat Lymphocytes in Plasma Cultures, (abstract) *Anat. Rec.*, vol. 52, p. 65.

Mihalik, P., 1932. Macrophages in Cultures of the Brain, (abstract) *Anat. Rec.*, vol. 52, p. 26.

Rich, A. R., and M. R. Lewis, 1932. The Nature of Allergy in Tuberculosis as Revealed by Tissue Culture Studies, *Johns Hopkins Hosp. Bull.*, vol. 50, pp. 115–131.

Lewis, W. H., 1934. On the Locomotion of the Polymorphonuclear Neutrophiles of the Rat in Autoplasma Cultures, *Johns Hopkins Hosp. Bull.*, vol. 55, pp. 273–279.

Chang, C., 1934. Some Studies on the Selachian Blood in vitro, (abstract) *Bull. Mt. Desert Isl. Biol. Lab.*, p. 27.

Lewis, W. H., 1935–1936. Pinocytosis, *Anat. Rec.*, vol. 64, suppl., p. 68.

Lewis, M. R., and E. G. Lichtenstein, 1936*a*. Breaking Down the Resistance of Albino Mice to the Transplantation of Tumors Induced by 1:2:5:6-dibenzanthracene in a Different Strain of Albino Mice, *Amer. Jour. Cancer*, vol. 27, pp. 246–256.

Lewis, M. R., and E. G. Lichtenstein, 1936*b*. Further Studies on the Breaking Down of the Resistance of Mice of One Strain to the Transplantation of Tumors from Mice of Another Strain, *Amer. Jour. Cancer*, vol. 28, pp. 746–751.

LEWIS, W. H., 1937. Pinocytosis by Malignant Cells, *Amer. Jour. Cancer*, vol. 29, pp. 666–679.

LEWIS, M. R., 1937*a*. Myeloid Hyperplasia Brought About in Mice by the Growth of Dibenzanthracene Tumors and Its Relation to the Transplantability of the Tumors into Mice of Alien Strains, *Amer. Jour. Cancer*, vol. 29, pp. 510–516.

LEWIS, M. R., 1937*b*. Myeloid Infiltrations Occurring in the Adrenals of Animals Bearing Certain Tumors, *Amer. Jour. Cancer*, vol. 30, pp. 95–101.

LEWIS, M. R., and E. G. LICHTENSTEIN, 1937. Studies on the Transplantability of Induced and Spontaneous Tumors Occurring in Mice of Pure Inbred Strains, *Growth*, vol. 1, pp. 375–383.

LEWIS, W. H., 1938. On the Rôle of a Superficial Layer in Division, Locomotion and Changes in Form of Cells, Internat. Cytol. Cong., Zürich, August 7–13.

EARLY DEVELOPMENT OF THE PRIMITIVE MESOBLAST IN EMBRYOS OF THE RHESUS MONKEY

CHESTER H. HEUSER

Department of Embryology
The Carnegie Institution of Washington

An adequate series of stages for the study of the first two and one half weeks of development in man and the other primates has always been lacking, but the available material shows that a striking peculiarity in the development of this order of mammals is the precocious formation of the exocœlomic space. The nature of the delicate network of fibrils permeating the albuminous fluid which occupies this space has not been understood. The strands have been named the "magma reticulare" and variously described as coagulation products of extracellular material or as connected with cells in some sort of way. The closely graded series of stages secured from our colony of rhesus monkeys permits the tracing of the various steps in development from the period of cleavage to term.

The first specimen obtained was an intact 22-day ovum. Upon opening it a thin veil-like membrane was found stretching loosely over the embryo to the margins of the chorion. Serial sections of this and additional specimens revealed that the membrane did not terminate at its peripheral contact with the chorion but continued to the abembryonic pole, thus forming a closed vesicle which occupied a large portion of the chorionic cavity. A report of this condition was given (Heuser, 1932).

The positions and relations of the structures within a chorion in which the wall of the vesicle is nearly complete are shown in a section of the 13-day ovum (plate 2, figure 12). The continued multiplication of the trophoblastic cells accompanied by a renewed spurt in the absorption of fluid results in a chorionic sac of enormous proportions in comparison to the size of the embryo proper. The latter consists of a thickened nearly circular germ-disk made up of large stratified radially arranged cells. Above it the amniotic cavity is closed in by a layer of cells merging with the covering trophoblast from which it has been derived, and on the ventral surface lies the yolk-sac. The remainder of the chorion is occupied by the exocœlomic vesicle.

After this brief reference to an intermediate stage of development, the examination of the material may be carried back to the time when the differentiation of tissues manifests itself in the blastocyst. The youngest blastocyst thus far obtained still possesses a zona pellucida which is more or less closely applied to the surface of the ovum. In general the egg consists of a thickened mass of cells, about 75 in all, continuous at its margins with a layer 1 cell thick which binds in the fluid-filled cavity of the ovum. Only a few of the cells, probably not more than 10 or 12, of the mass and about 7 slightly flattened cells (primitive entoderm) which have arranged themselves on its ventral or inner surface will take part in the formation of the embryo proper. The remainder of the cells form extra-embryonic or temporary portions of the egg. Thus out of the 75 cells making up the inner cell mass the vast majority, about 56, probably take no part in the building up of the embryo but represent a group of cells which will continue to increase rapidly in number in preparation for attachment to the maternal mucosa. In a slightly older blastocyst, although all types of cells have increased in number, the enlargement of the egg is due chiefly to growth of the trophoblast and further absorption of fluid. Photographs of these two specimens, no. C 614 and no. C 522, are reproduced on plates 1 and 2 in the paper by Streeter in this volume.

None of the older blastocysts are found free in the uterine cavity. The trophoblast first grows thicker in a zone around the embryo and unites with the endometrium. Two areas of contact are shown in plate 1, figure 1, the space between them being a part of the uterine cavity. Cells delaminating from the inner surface of the thickened trophoblast seem to be continuous with the primitive entoderm beneath the inner cell mass. However, it appears that we are dealing with two types of cells. Those in contact with the embryonic mass are more darkly staining and stand out in sharp contrast to all others: this is especially noticeable in slightly older stages, as shown in plate 1, figures 3–5.

The chorion is observed in a semicollapsed state in every case during the 10th and 11th days (plate 1, figures 3, 6; plate 2, figures 7–9). The trophoblast is constantly growing but for a time the absorption of fluid does not keep pace with the spread of tissue. In older stages the distended condition of the sac is restored. The same phenomenon was noted for a certain period in the development of the pig (Heuser and Streeter, 1929).

Primitive mesodermal cells delaminating from the trophoblast at the distal side of the chorion are seen for the first time in the 11-day embryo (plate 1, figure 6). The cells are not yet united together into a complete membrane. A second 11-day embryo (plate 2, figures 7–9), however, shows no mesodermal cells in the side of the chorion where the secondary placenta will later develop, although it is older in respect to the condition of the embryo itself.

This particular specimen is an important one. The primitive entoderm up to this time has remained in a relatively quiescent state and advanced no farther than a single layer of undifferentiated cells. These now multiply at a more rapid rate and begin to separate into layers. Nuclear and cytoplasmic strands shown in the three successive sections of the embryo (plate 2, figures 7–9) are parts of an internal layer in process of separation from cells which remain in contact with the germ-disk. The latter are destined to form the yolk-sac, which makes a belated appearance within a few hours. The internal layer retains continuity with the primitive mesoblastic sheet lateral to the germ-disk to complete the part of the exo-cœlomic membrane in this region.

The process of separation of these structures is nearly complete in the 12-day embryo (plate 2, figure 10). The membrane is continuous and except for a few bridging cells stands off at some distance from the yolk-sac.

During the following day mesodermal cells bud off the inner surface of the entire chorion and join together to form a complete membrane lining the ovum (plate 2, figure 11). The exocœlomic vesicle so formed occupies practically the whole of the chorionic cavity. The part of the membrane below the embryo during the 13th day is usually nearer the yolk-sac than as shown in figure 11. The specimen shown in plate 3, figure 13 is more typical for this period of development.

Up until now, except in a small zone around the embryo itself, there is little if any additional mesoblastic tissue. Progressively the trophoblast thickens, especially at the anti-embryonic pole where the secondary placenta will develop, and gives rise to mesenchyma, which builds up between it and the exocœlomic vesicle.

During the next four or five days the amniotic cavity increases rapidly in size. This change along with the growth of the germ-disk and yolk-sac causes an increasing bulge of the exocœlomic membrane into the chorionic cavity. A representative of this stage of development is the 17-day embryo. The section shown in plate 3, figure 15, passes through the middle of the embryo. It will be noted that a few cells have made their appearance between the germ-disk and the gut entoderm. These cells, located within the embryo, are embryonic (sometimes less appropriately named secondary) mesoblast and have a different origin from that of the primitive mesoblast. At the edges of the germ-disk, however, the two types of mesoblast blend and probably some cells of the primitive type persist. The yolk-sac, although considerably larger than in younger stages, has not grown as rapidly as the surrounding structures and its diameter is still much less than that of the germ-disk.

Between the 17th and 22d days the differential rate of growth favors the yolk-sac so that it soon reaches beyond the confines of the embryo itself and protrudes far out toward the center of the chorion. Two embryos collected on the 22d day possess one pair and eight pairs of somites respectively. The younger one, plate 3, figure 16, was cut open after fixation through the thin zone between the primary and secondary placentas without disturbing any of the internal structures, and was photographed to show the condition of the exocœlomic membrane. Gaps are seen for the first time in the membrane both near the edges of the chorion and above the bulging yolk-sac. The perforations increase rapidly in number until the membrane becomes transformed into a fenestrated sheet. It then changes over into a meshwork of broad and narrow bands, veil-like areas persisting for a time at angles in the mesh. Such veil-like remnants are shown in the photograph of the 8-somite embryos (plate 3, figure 17).

Owing to the growth of the embryo the part of the membrane stretching over the yolk-sac gradually bulges out farther and farther into the chorion and approaches its counterpart on the opposite wall. The exocœlomic vesicle in this way is reduced and becomes obliterated on about the 23d day of gestation. The cavity of the chorion thus becomes bridged by a network of mesenchymal strands, as seen in plate 3, figure 18, which are attached to the inner lining of the chorion and to the external coats of the amnion and yolk-sac. The latter structure increases in size only slightly after this time, whereas the rapidly expanding amnion appropriates the space previously occupied by the network of primitive mesoblast, which is finally reduced to a membrane of fibrils between the amnion and chorion.

BIBLIOGRAPHY

Brewer, J. I., 1938. A Human Embryo in the Bilaminar Blastodisc Stage. Carnegie Inst. Wash. Pub. No. 496, *Contr. to Embryol.*, vol. 27, pp. 85–93.

Bryce, T. H., 1924. Observations on the Early Development of the Human Embryo, *Trans. Roy. Soc. Edinburgh*, vol. 53, pp. 533–567.

Florian, J., 1933. The Early Development of Man, with Special Reference to the Development of the Mesoderm and Cloacal Membrane, *Jour. Anat.*, vol. 67, pp. 263–276.

Heuser, C. H., and G. L. Streeter, 1929. Early Stages in the Development of Pig Embryos, from the Period of Initial Cleavage to the Time of the Appearance of Limb-Buds. Carnegie Inst. Wash. Pub. No. 394, *Contr. to Embryol.*, vol. 20, pp. 1–30.

——— 1932. An Intrachorionic Mesothelial Membrane in Young Stages of the Monkey (*Macacus rhesus*), *Anat. Rec.*, vol. 52, suppl. 15.

Hill, J. P., 1932. (Croonian Lecture) The Developmental History of the Primates, *Phil. Trans. Roy. Soc. London*, Ser. B, vol. 221, pp. 45–178.

Selenka, E., 1900. Entwickelung des Gibbon (Hylobates und Siamanga), Fortsetzung. Studien über Entwickelungsgeschichte der Tiere, No. 8, Menschenaffen, pt. 3, pp. 173–208.

Wislocki, G. B., and G. L. Streeter, 1938. On the Placentation of the Macaque (*Macaca mulatta*), from the Time of Implantation until the Formation of the Definitive Placenta. Carnegie Inst. Wash. Pub. No. 496, *Contr. to Embryol.*, vol. 27, pp. 1–66.

PLATE 1

Photographs of sections of monkey embryos:

FIG. 1. The section shows the structure of the blastocyst at the time of implantation. The trophoblast is fused with the uterine mucosa in two places, one on either side of the inner cell mass; space above it is a part of the uterine lumen.

8-day embryo, no. C 610. Section 3–4–4, × 150

FIG. 2. This adjoining section of the same embryo is more highly magnified to show features of the various kinds of cells. The layer of primitive entoderm is sharply marked off from the embryonic cells above it.

8-day embryo, no. C 610. Section 3–4–3, × 375

FIG. 3. The contrast between primitive entoderm and formative cells is further emphasized in this section. The absorption of fluid has not kept pace with the growth of the trophoblast and the chorion appears collapsed.

10-day embryo, no. C 548. Section 1–9–7, × 225

FIG. 4. The radially arranged formative cells have established a germ-disk. The space between it and the covering trophoblast is the primitive amniotic cavity. On the lower surface of the disk the primitive entoderm remains as a single layer of darkly staining cells.

10-day embryo, no. C 496. Section 4–6–10, × 225

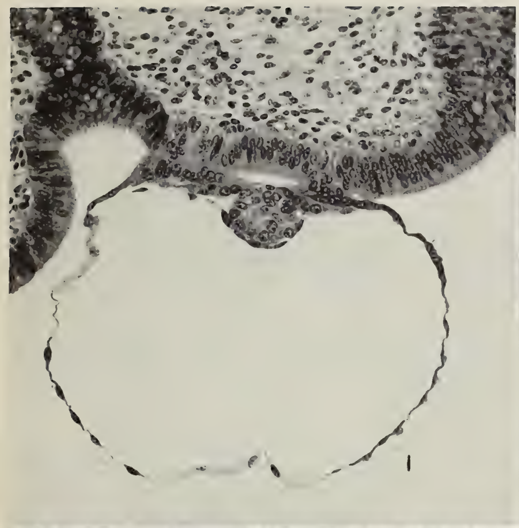
FIG. 5. The primitive entoderm is still a single layer of cells, although in some of them the cytoplasm has increased.

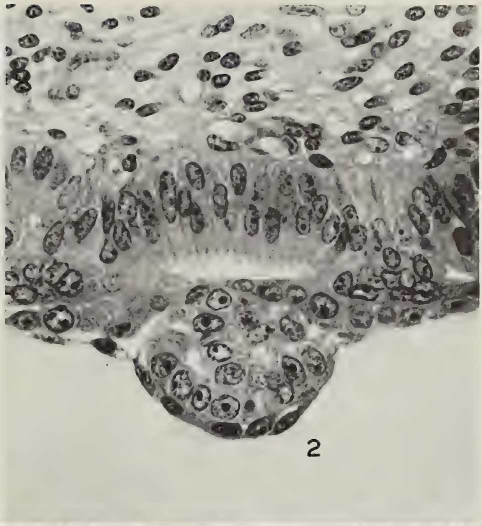
10½-day embryo, no. C 596. Section 4–2–10, × 225

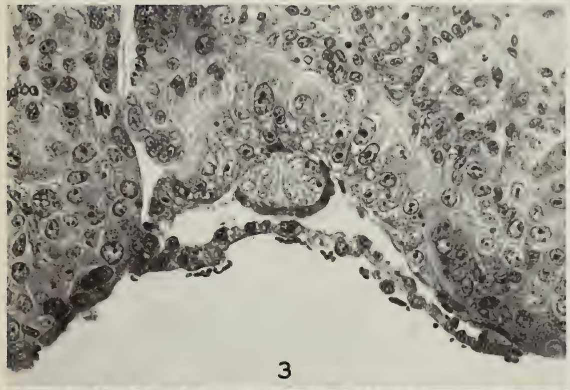
FIG. 6. Primitive mesoblast is now seen in all regions of the trophoblast but the membrane is not quite complete. The chorion is still in a semicollapsed state. The primitive entoderm is beginning to give rise to mesoblast.

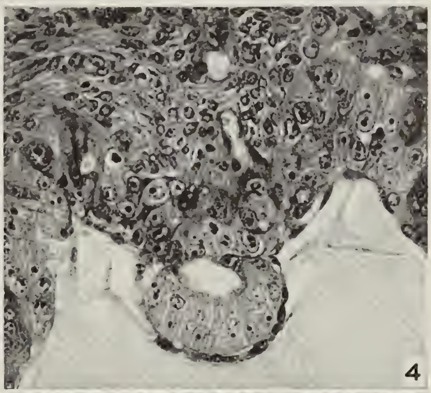
11-day embryo, no. C 602. Section 3–2–9, × 225

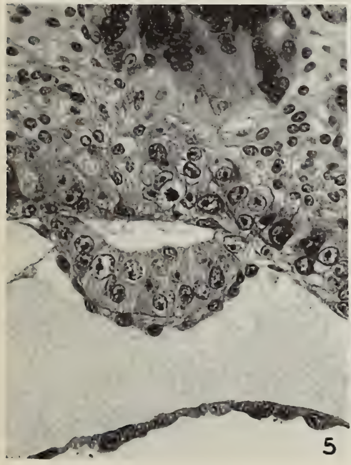
[HEUSER] PLATE 1


1


2


3


4


5

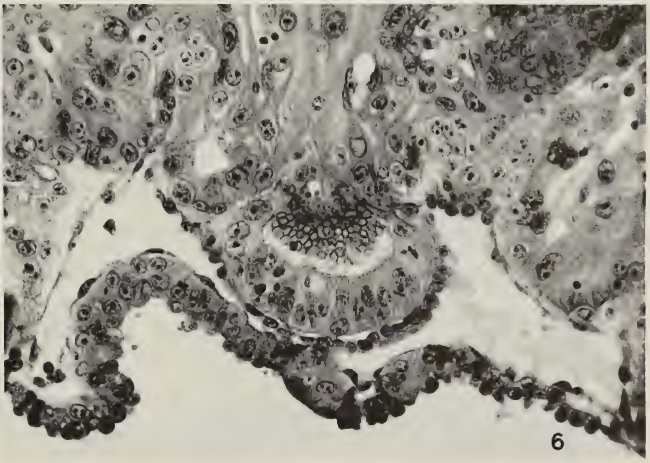

6

PLATE 2

The first three photographs show sections in serial order through an embryo at a key stage of development. Marked activity is indicated in the primitive entoderm, which is separating into mesoblast and cells which soon will give rise to the yolk-sac.

FIG. 7. 11-day embryo, no. C 524. Section 3–4–11, × 375

FIG. 8. 11-day embryo, no. C 524. Section 3–4–12, × 375

FIG. 9. 11-day embryo, no. C 524. Section 3–5–1, × 375

FIG. 10. The mesodermal membrane is now a continuous layer and almost completely separated from the yolk-sac. Strands connecting the two structures are still present.

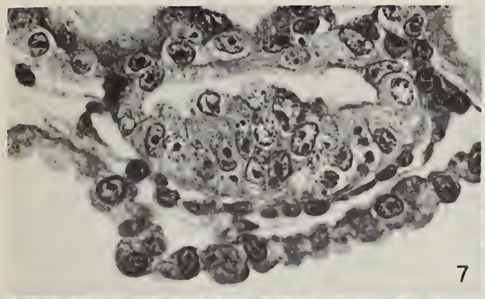
12-day embryo, no. C 530. Section 3–2–7, × 225

FIG. 11. The chorion has become distended again with fluid. The exocœlomic vesicle is completely formed.

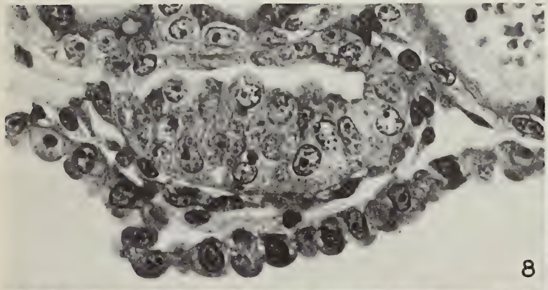
13-day embryo, no. C 559. Section 3–8–6, × 75

FIG. 12. The mesodermal membrane is nearly complete. The section shows the marked difference in thickness of the yolk-sac and gut entoderm.

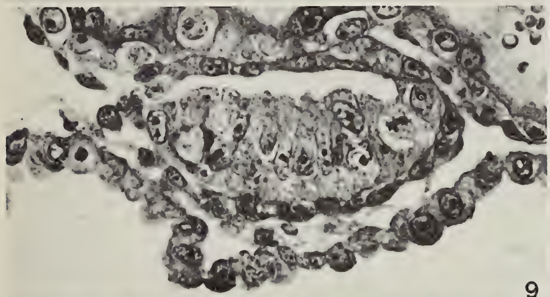
13-day embryo, no. C 609. Section 4–1–5, × 120



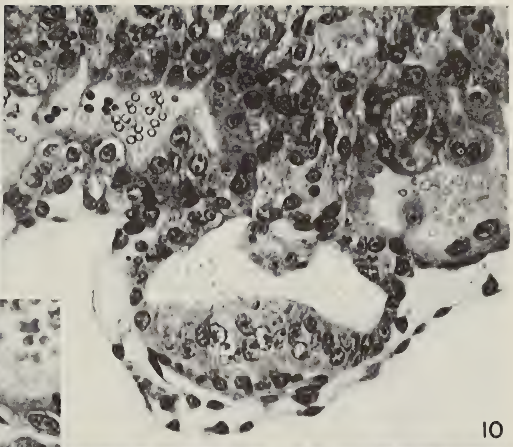
7



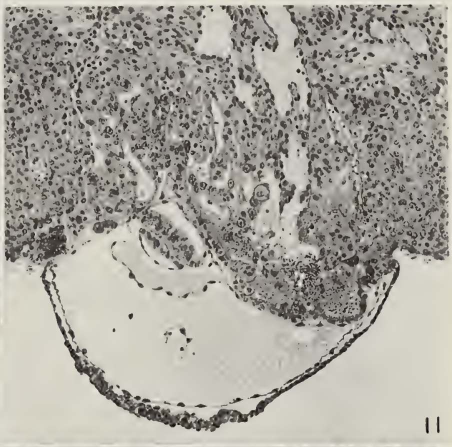
8



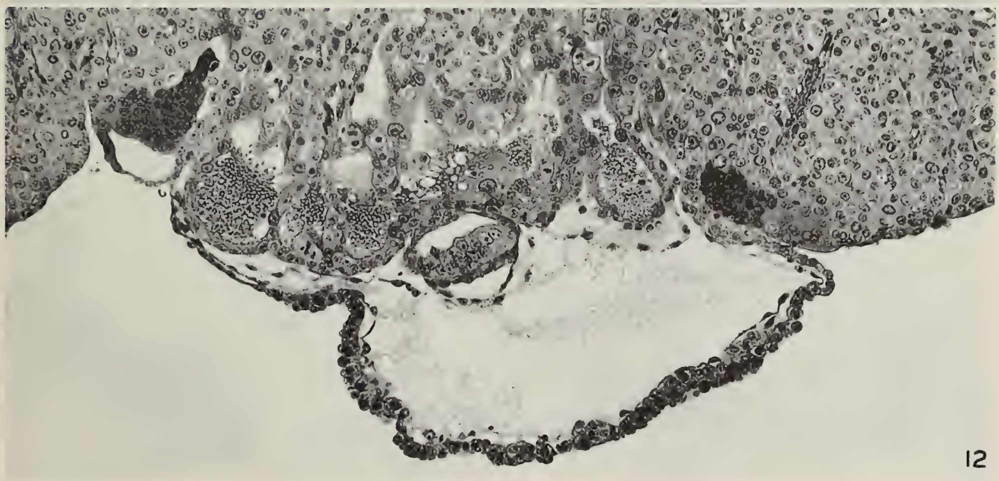
9



10



11



12

PLATE 3

FIG. 13. The section shows the distended chorion completely lined with the thin mesothelial wall of the exocœlomic vesicle. Numerous mesenchymal cells occur near the embryo.

13-day embryo, no. C 466. Section 4–6–2, × 37.

FIG. 14. The next section of the same embryo shows a few free cells lateral to the yolk-sac. These primary mesodermal cells will take part in the formation of the external coat of the yolk-sac.

13-day embryo, no. C 466. Section 4–6–1, × 225

FIG. 15. The loose character of the primary mesoblast around the embryo is shown in this section. Embryonic mesoblast is also growing between the germ-disk and gut entoderm. The two types of mesoderm blend.

17-day embryo, no. C 457. Section 1–7–7, × 112

FIG. 16. The chorion was cut open and photographed with the embryo *in situ* to show the perforations which have appeared in the exocœlomic membrane.

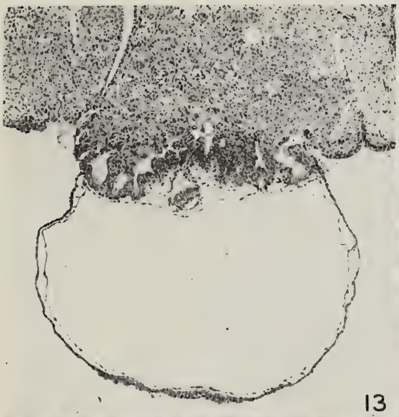
22-day embryo, no. C 475. Chorion × 6

FIG. 17. The embryo still lies with its dorsal surface toward the primary placenta and the yolk-sac protruding into the chorion. The membrane has been transformed into a meshwork of broad and narrow bands within chorion cavity.

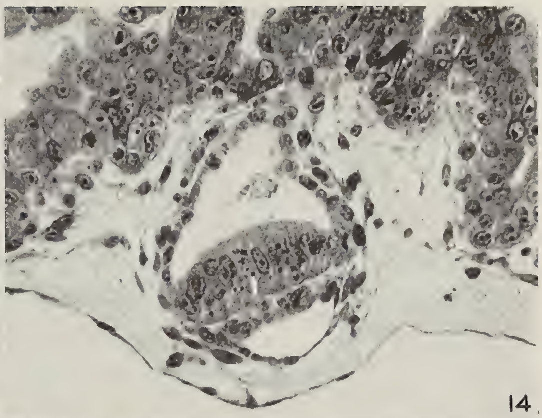
22-day embryo, no. C 474. × 6

FIG. 18. The body-stalk has increased in length permitting the embryo to rotate on its side. The yolk-sac is shown on the right side of the photograph. A network of fibrils extending between the yolk-sac amnion and the chorion now occupies the chorionic space.

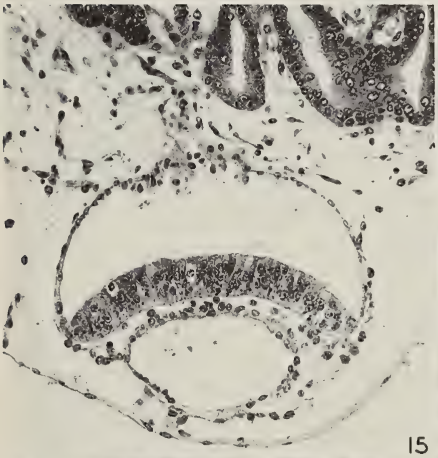
31-day embryo, no. C 499. × 4



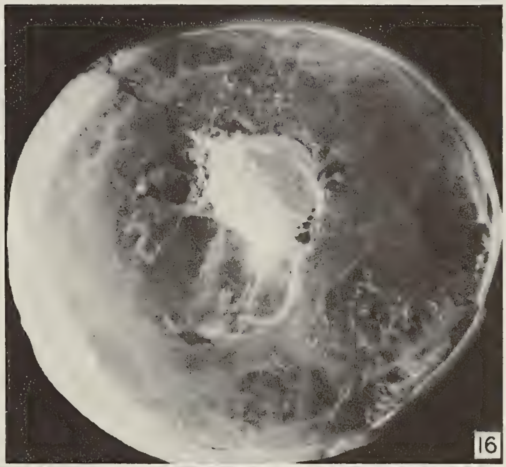
13



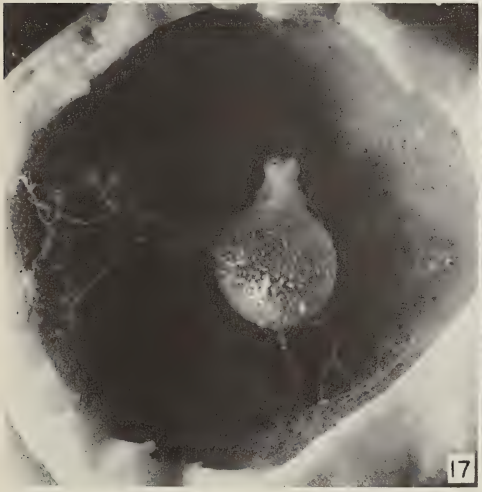
14



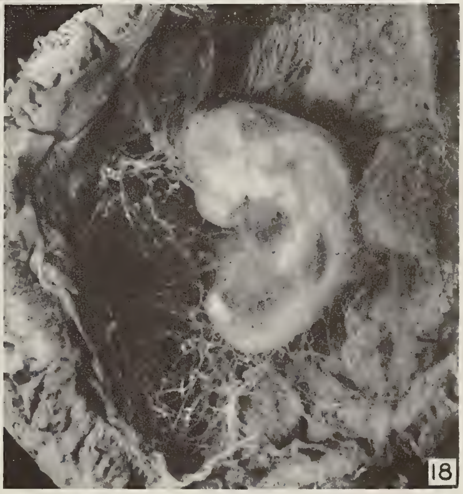
15



16



17



18

ATTEMPTS TO INDUCE OVULATION IN THE ADULT RHESUS MONKEY

CARL G. HARTMAN

Department of Embryology
The Carnegie Institution of Washington

It has been shown possible to induce ovulation in a variety of vertebrates by the use of ovary-stimulating (gonadotropic) substances from various sources: anterior pituitary gland, blood serum of pregnant mares, urine of pregnant women or of women who have passed the menopause. The exigencies of space preclude an adequate discussion of this subject here, but a few statements taken from the writer's discussion of the subject in Allen's *Sex and Endocrine Glands* (new edition, 1938) are here added by way of introduction.

If an animal is already in estrus, presumably, therefore, possessing in its ovaries one or several mature follicles, ovulation may be precipitated by the intravenous injection of any of the gonadotropic substances, as, for example, in the rabbit test for pregnancy (Friedman). The method has an obvious advantage in controlling the time of ovulation, as in the mare, rendering mating and artificial insemination more certain of favorable results.

In juvenile animals the first estrus and the first ovulation may be brought on prematurely by the administration of an appropriate gonadotropic substance. This fact is, of course, of no especial interest to the animal husbandman, for the breeding of an animal at too early an age is manifestly deleterious. Of practical importance, however, is the fact that in species like the sheep and the goat which have normally a limited breeding season the animals may be made to breed the year around. Thus, for example, for persons allergic to cow's milk and therefore dependent on goat's milk a more continuous supply of the latter will be possible when kidding is more evenly distributed throughout the year.

Means seem at hand, therefore, to speed reproduction in mammals as the horticulturist "forces" maturation in plants.

Therapeutic Application to the Rhesus Monkey

Since the organization of the Carnegie colony of rhesus monkeys, begun in 1925 under the stimulus of Dr. Merriam and Dr. Streeter, a large amount of data has been accumulated on the breeding performance of the monkey under fairly favorable conditions. According to the writer's figures published in 1937, many menstrual cycles are experienced without the occurrence of ovulation even by the best females. The failure to ovulate, if continued, leads of course to sterility. Just such cases have been encountered in numbers in the colony. And since gynecologists are now finding the counterpart of this deficiency in women, and since "cures" of this condition have been reported in clinical literature, the monkey colony seemed favorable for a study of the action of the "curative" extracts under conditions that could be adequately controlled.

All the experiments were performed in the breeding season of the rhesus monkey (October to April or May), when it is "normal" for a monkey to ovulate in each and every menstrual cycle. It would be logical on another occasion to repeat the present experiment in the summer months, when rhesus females seldom ovulate; to date this has not been found practical. In the present series of 104 experiments mostly non-ovulating females were selected; in others attempts were made to control the *time* of ovulation by the introduction of the extracts in a single large dose given intravenously.

The chief extract employed was Gonadogen, prepared from pregnant mares' serum by the Upjohn Company. Gamone Squibb and Follutein Squibb as well as purified follicle-stimulating hormone (F. S. H.) prepared in the laboratory of Dr. F. L. Hisaw were also used. Thanks are due these organizations for their generous cooperation.

The details of the experiments are omitted with the exception of several protocols which serve to illustrate the nature of the evidence on which the conclusions are based. The diagnosis of results was made in part by the writer's method of bimanual palpation, frequently checked by laparotomy, at which time the ovaries and uterus could be inspected. Frequently an ovary was removed and the uterus biopsied, with or without recovery of an ovum or implanted embryo. Some animals were killed at the end of an experiment and the organs studied histologically.

Four types of results are recorded: (1) there was little or no response to the injections; (2) large normal follicles were produced; (3) the ovary was overstimulated and the follicles palpably injured; (4) the desired ovulation occurred.

1. *Animals showing little or no response to the injected hormones*

In about one-third of the cases (32 out of 104) the ovaries showed little or no response, although the dosages used were comparable to those effective in other animals or indeed in the same animal at other times, as appeared from the records of a dozen animals used more than once.

When the cases of refractoriness now under discussion were compared with the responsive cases listed in section 4 below, the former proved to be practically all non-ovulators, the latter mostly ovulators when the experiments were begun. In other words, if an animal is refractory to its own pituitary hormones it is likely to remain refractory to moderate amounts of introduced extracts.

Of the 32 complete failures to elicit any visible growth of follicles with the follicle-stimulating hormone, 19 cases find some sort of explanation: (1) Eight animals, though bleeding periodically, were palpably sick. No. 535 had a skin affection but seemed otherwise healthy. Nos. 267 and 424, once excellent animals, became refractory to Gonadogen in later stages of tuberculosis. (2) Nine females had not yet become "acclimated," that is, had not yet recovered from the debilitating journey from India. (3) Two females (nos. 57 and 114) were over eleven years old, which, however, is some years removed from the menopause, a conclusion based on our experience with females nos. 5 and 7. These were still menstruating though not ovulating at eighteen and nineteen years of age.

For the remaining 13 complete failures no "explanation" suggests itself. All of them had been ovulating more or less normally but had for unknown causes lapsed to physiological states somewhat below perfection. The vagaries of responsiveness to hormones in an animal may be illustrated by the behavior of monkey no. 430, as follows:

HISTORY OF No. 430

EXP. 1. Non-ovulatory cycles in 1936–1937 were 56, 20, 22, 20, 28 days in length. Menses, February 23, 1937; injection of 5 U. U. Gonadogen s. c. daily on days 5 to 9, 10 units i. v. on day 10; on March 9, day 15, laparotomy; ovary removed contained large, palpably injured follicles with abnormal stigmata (figure 4). The ovaries had been overstimulated (figure 7). Next bleeding March 17–24.

Exp. 2. Repetition of experiment 1, this time with no palpable increase in ovary or uterus nor in vaginal desquamation. Since the latter constitutes an extremely delicate sign of ovarian activity, we may assume that there was no stimulation whatever.

Exp. 3. Ovulation in October 1937, after injection of 1 U. U. Gonadogen daily on days 3–14. In the succeeding cycle ovulation took place spontaneously, conception occurred, and a 12-day ovum was recovered by hysterotomy. The January 1938 cycle of 36 days was also ovulatory. Bleeding on January 27.

Exp. 4. Since the animal's sex skin grew paler at this time and the uterus remained small (two indications of lower hormone level), she was injected for 9 days with the relatively large dosage of 12 U. U. Gonadogen daily; followed by 30 units i. v. on February 12. She was killed on February 14. The ovary (figure 3) was found to contain corpora albicantia one and two months old respectively, but no follicle over 0.5 mm in diameter. The uterine glands (figure 6) contained numerous mitotic figures but the vaginal mucosa none; hence the response was slight.

It seemed, therefore, that when this animal was refractory to her own anterior pituitary hormones she was refractory also to the hormones administered to her. It is incidentally to be noted also that the ovary fully recovered from the precocious stimulation.

The 32 experiments thus far discussed included failures that found an explanation for the refractoriness of the animals in their physical condition. The remaining 13 females, however, were apparently in as good condition as the more responsive animals to be mentioned in the following sections.

2. *Normal follicular enlargement without ovulation*

This series of experiments includes 7 cases, including No. 430^{I} described above. Seven times a large ovarian follicle of normal appearance, usually around 4 mm in diameter, was discovered in the course of the experiment. How many of these were sufficiently normal eventually to rupture is a matter of conjecture. In one case (No. 451^{II}) a follicle seen at laparotomy did actually rupture later, but no such positive diagnosis was made in other cases. Perhaps this case should be added to the "successes" treated in section 4.

3. *Diffuse overstimulation of the ovaries*

Sixteen experiments resulted in ovaries greatly overstimulated. Most often several or many follicles were involved; in only 3 cases a single follicle. Twice the effect was unilateral, the right ovary only responding in riotous fashion.

When a single follicle responds (probably because it already has a start on its fellows when the injections began), it reaches a size never seen in other cases. But even such follicles can hardly be termed cystic. They are not excessive in size, 8 mm in diameter as a maximum, and structurally they have little in common with spontaneously occurring cysts, of which the writer has seen more than a dozen. Monkey no. 427, indeed, though she possessed an ovarian cyst which reached the size of a hickory nut in December 1936 and January 1937 (menses December 11, January 4, February 3), developed only enlarged follicles (figure 8) and no cyst as the result of injections in February. The ovary of no. 460 is of especial interest, for it accidentally revealed the disrupted character of the blood vessels characteristic of such cases. The animal had been killed by intravenous injection of air, some of which found its way to the ovarian blood vessels and thence into the large graafian follicle.

In several instances the ovaries contained one or two large follicles, many small ones. But the usual picture is that of diffuse, general overstimulation of the ovaries, so that in sections one sees little ovarian substance other than follicle walls (figure 1). These are thicker than one would expect at such stages owing to the undue stimulation of the granulosa (figure 2). This is rich in mitotic figures, even after chromatolytic degeneration of the cells has already begun at the inner surface of the membrane. It is the rapid progress of degenerative processes which limits the size of the follicles and renders them unsuitable for rupture. Ovulation, or the opening of the follicle for the release of the ovum, is the result of an orderly differentiation of the follicle wall (Hartman in Allen's *Sex and Endocrine Glands*) and is extremely sensitive to injurious stimuli. Thus, for example, at laparotomy of no. 431, on day 9 of the cycle, an apparently normal follicle in the pre-ovulatory stage of growth was encountered. Whether ovulation would have occured if the animal had remained untreated one can only guess, but injection of 20 U. U. of Gonadogen was in this case not followed by dehiscence of the follicle, but by atresia. In another case (no. 442^{III}), however, a similar treatment in the presence of a similar follicle led to its rupture and its metamorphosis into a normal corpus luteum.

One of the indications of injury to a follicle is the peculiar shape and color of the stigma over the distended follicle as seen in the living state (figure 7). Instead of being thin, clear, and translucent it is elongated, opaque, and red. The region is only a little thinner than the cortex generally and therefore not fitted to rupture. It should be stated, however, that similar abnormal stigmata were twice recorded in untreated animals, which again illustrates the fact that phenomena which are the results of our experimental procedures are often manifested, under abnormal conditions, in untreated animals, both in the laboratory and in the wild state.

It is thus seen that doses which failed to have the slightest effect on the animals listed in section 1 played havoc with the ovaries of the animals listed in this section. Two animals (nos. 205 and 430), refractory at one time, responded excessively the next.

4. *Ovulation concomitant with injection*

In this section 28 cases are included in which ovulation accompanied treatment with gonadotropic hormones. It should be emphasized at the outset, however, that in this statement cause and effect is not necessarily implied. Indeed, in the first 5 cases on the list it was possible to prove, by adequate control, that ovulation actually *preceded* the administration of the hormone! Because of the possibility of making examinations by laparotomy or even autopsy at will in the monkey, only the last seven "cures" can be "claimed" for the treatment employed. In between these two extremes are 16 doubtful cases that may have ovulated spontaneously. Several details will illustrate the nature of the evidence.

Ovulation occurring before injection. The evidence in the five cases involved is as follows:

In no. 395 a 5- or 6-day corpus luteum was removed 2 days after the injection. A similar finding holds for no. 432.

In no. 431^{III} a 9-day vesicle (C 613) was recovered on January 24, which fixes the date of ovulation as prior to the injection of the "ovulating" dose of Gonadogen.

As palpation of no. 466^{II} had failed to recognize ovulation by day 11 of the cycle, 10 U. U. of Gonadogen was injected on October 14. Recovery of a 13-day unplanted ovum on October 25 proved palpation to have been faulty, and the injection futile.

The ovum and the corpus luteum recovered from no. 433 on the day after she was given 30 units of Gonadogen were both nearer 3 days than 24 hours old.

Animals that would possibly have ovulated without treatment. Of such animals 12 were ovulating both before and after treatment, 3 after but not before. No. 237 had been amenorrhic for four months preceding the treatment; ovulation was determined by palpation 24 hours after injection, a diagnosis corroborated by laparotomy three days later. Surely one would be inclined to add this instance to the list of "successes"; yet the animal ovulated spontaneously and conceived the very next cycle. No. 464^{I} presents a somewhat similar history.

Ovulation with reasonable certainty due to injection. Fair to excellent evidence is offered in the results of 8 experiments that ovulation was actually to be ascribed to the administered reagent. Of these 4 were not mated (nos. 330, 424^{I}, 265^{II}, 442^{II}), 3 were mated but did not conceive (nos. 444^{I}, 444^{II}, 464^{I}), and one conceived (No. 442^{IV}). The last furnished a beautiful ovum in process of attachment (figures 12 and 13). All the facts surrounding this last case make it practically certain that this ovum would not have been secured without the use of the follicle-stimulating hormone. The corpus luteum is shown in figure 11, the early pregravid endometrium to the surface of which the ovum was beginning to attach itself in figure 10.

Summary and Conclusions

Since it is the growing conviction among gynecologists that many cases of sterility in women are due to the failure of ovulation, it seemed desirable to test the therapeutic value of the new follicle-stimulating substances in the rhesus monkey. This experimental animal is particularly adapted to such a study since the incidence of non-ovulatory menstrual cycles is high and the effect of medication can be adequately controlled, first by a preliminary study of the animal before the treatment and later by a thorough examination of the organs removed at laparotomy or at autopsy.

The reagent mostly employed in the 104 experiments here recorded was the Upjohn Company's very potent product Gonadogen, made from the serum of pregnant mares and discovered by Cole and coworkers to contain a high percentage of follicle-stimulating hormone (F. S. H.) as compared with the luteinizer (L. H.).

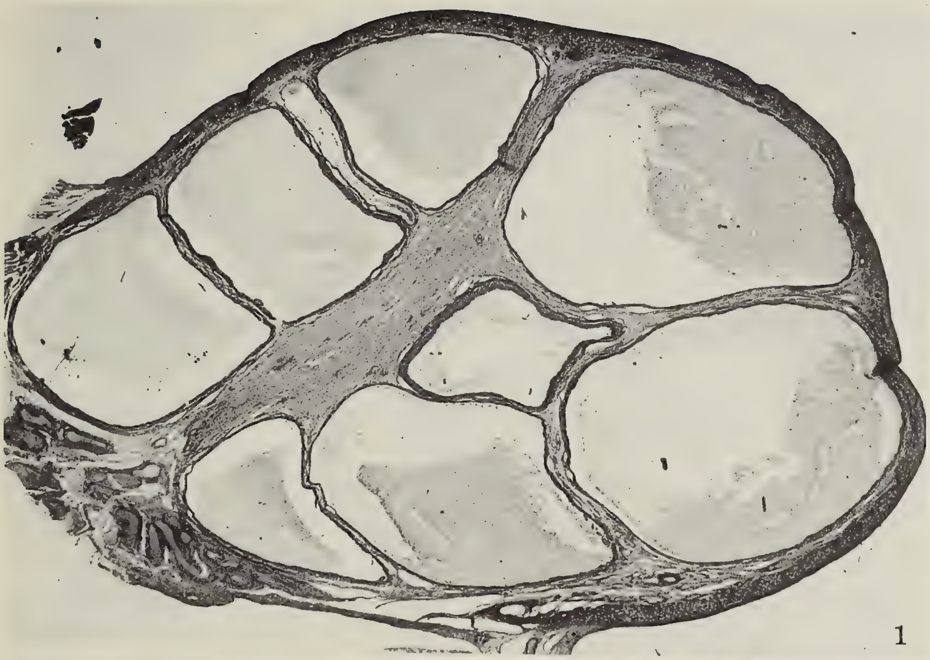
The results were far from promising. In 32 cases the animals remained perfectly refractory and in only half of these was a poor general bodily condition associated with the refractoriness of the ovaries. The doses were quite comparable to those which in other females and at other times in the same animal proved adequate or even much too potent. In general it may be said that if the gonads are refractory to the animal's own anterior pituitary they are likely to prove refractory to injected extracts also.

Overstimulation of one or usually many graafian follicles was a common result. Their growth proceeds too rapidly, producing large but injured follicles incapable of rupturing. The rapid growth of the follicles renders it impossible for the uterus to keep pace; hence the endometrium may be inadequately prepared for the reception of the fertilized egg. There is some indication in the material that under the artificial stimulus, even the production of ovulation does not guarantee an endocrine condition favorable to implantation, for in some cases the menses recurred very soon after ovulation, in No. 464[1] as early as the sixth day.

The rapid regression of injured follicles accounts for the shortening of the menstrual cycle after treatment. This happened even where the effect of the injections was not grossly recognizable. The average cycle length in 19 cases was 22 days; in 13 cycles where ovulation occurred it was 26 days, though the latter series contained three cycles of 18, 19, and 21 days respectively.

To summarize, then, we find a mere 7 per cent of clear-cut successes—the production of ovulations which would not have taken place if no treatment had been given. Is treatment with potent endocrines justifiable under these circumstances, especially since the hormones often damage the ovaries? The answer is that the damage is not permanent so far as the present material shows, for recovery is prompt and complete, the ovaries resuming perfectly normal functioning spontaneously.

The explanation of the failures lies in the difficulty of estimating the dosage, for the amount of hormone produced by the experimental animal's own pituitary varies greatly; it has been shown that a dosage inadequate to raise the level of circulating hormone to the ovulation point in one case is sufficient temporarily to disrupt the follicular apparatus in another. One should expect to encounter the same uncertainty in treating women.



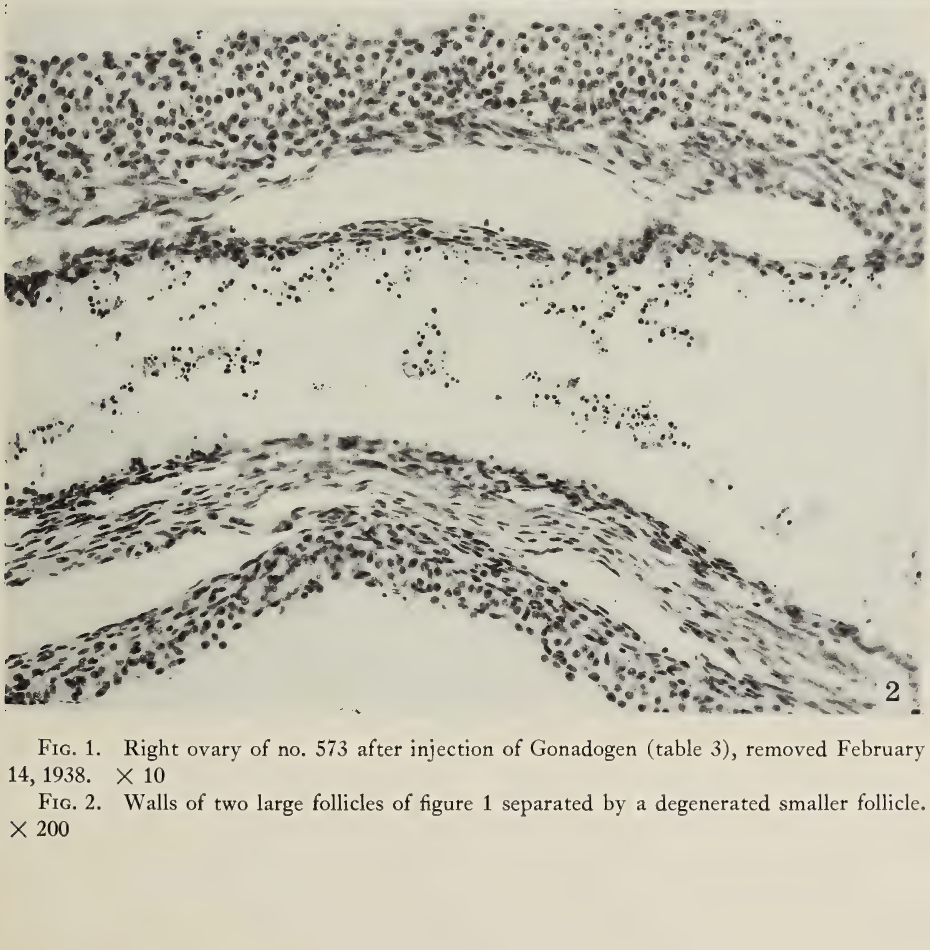


FIG. 1. Right ovary of no. 573 after injection of Gonadogen (table 3), removed February 14, 1938. × 10

FIG. 2. Walls of two large follicles of figure 1 separated by a degenerated smaller follicle. × 200

Plate 2

Fig. 3. Left ovary of no. 430 removed February 14, 1938, showing normal condition, but non-effect of injection of large amount of Gonadogen. The corpora lutea shown are one and two months old. × 10

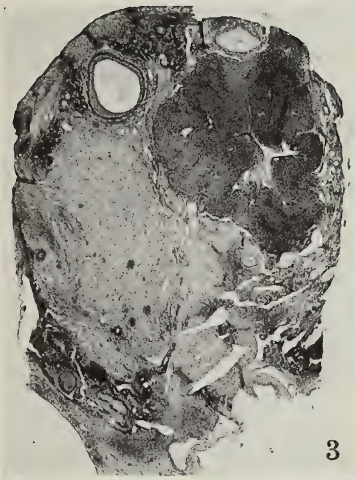
Fig. 4. Right ovary of same animal showing response to Gonadogen, March 9, 1937 (cf. figure 7). × 10

Fig. 5. Right ovary of no. 442 showing stigma of newly ruptured follicle, November 30, 1937. × 4. Section shown in figure 9, plate 3.

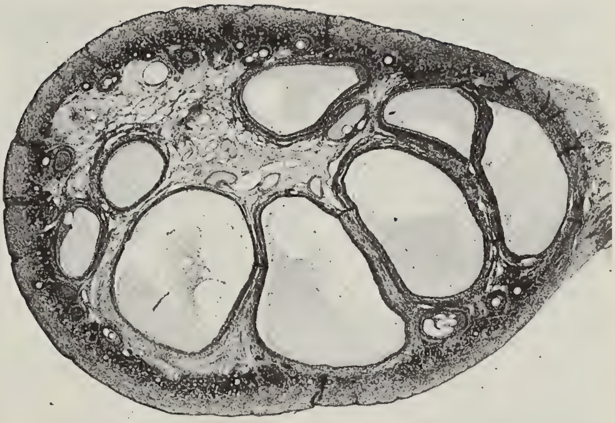
Fig. 6. Section of uterus of no. 430, of which corresponding ovary is shown in figure 3. × 100

Fig. 7. Right ovary of no. 430 shown in section in figure 4. The dark streak across the ovary is an aberrant "stigma" characteristic of overstimulation. × 4

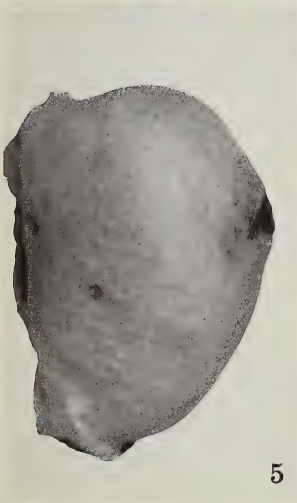
Fig. 8. Left ovary of no. 427 (April 16, 1937), split and photographed in alcohol after fixation. Numerous follicles have been stimulated to growth by Gonadogen. × 4



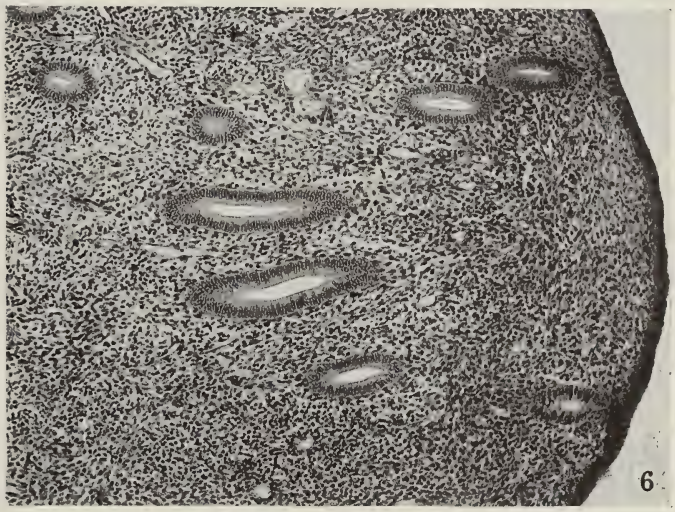
3



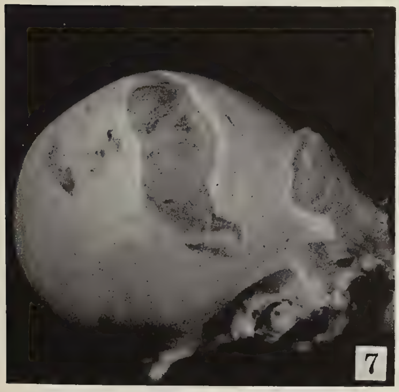
4



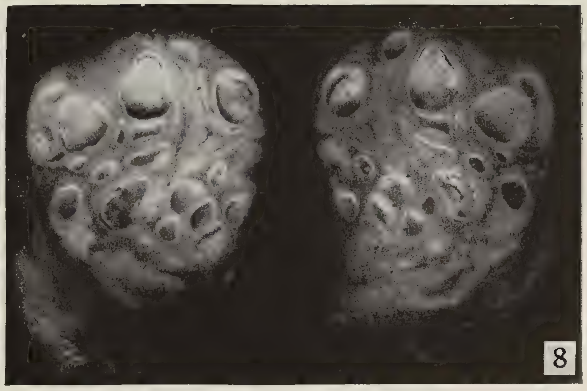
5



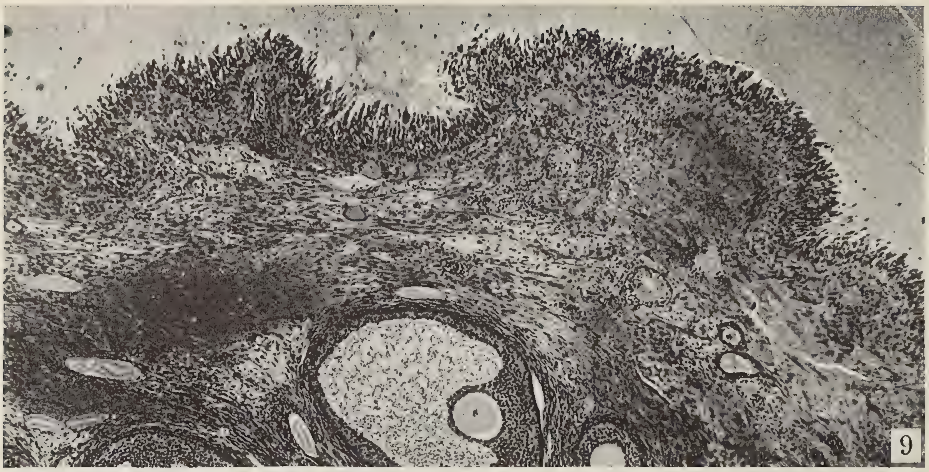
6

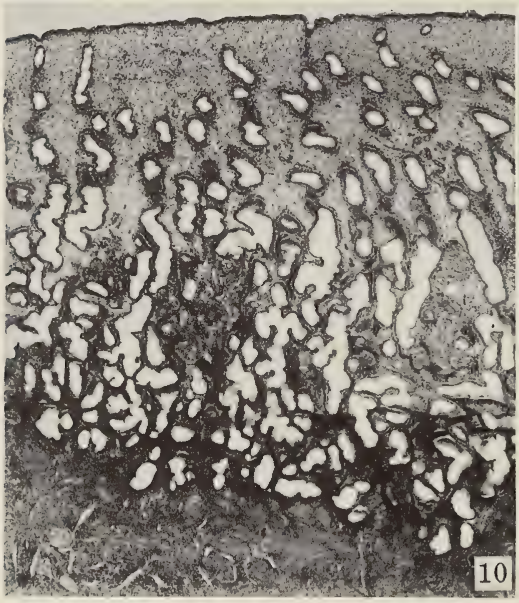


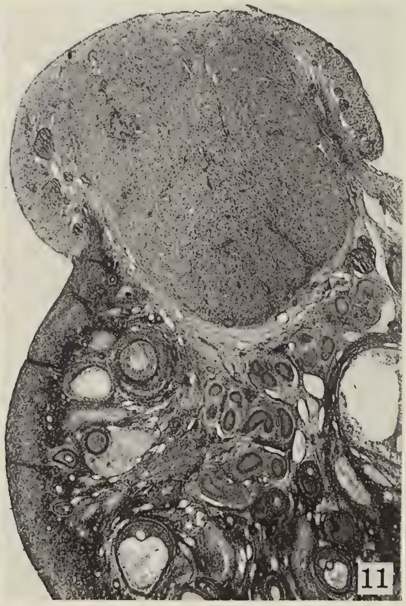
7

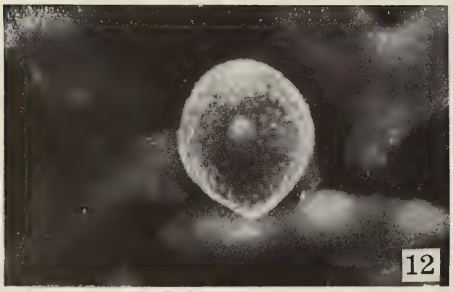


8









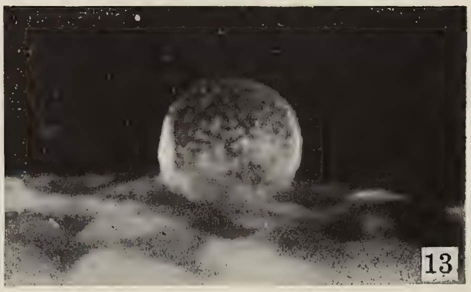


FIG. 9. Portion of freshly ruptured follicle of no. 442, shown in figure 5, November 30, 1937. × 50

FIG. 10. Endometrium of no. 442, January 19, 1938. The pregravid condition was controlled by the corpus luteum shown in figure 11. × 15

FIG. 11. Corpus luteum of no. 442, January 19, 1938. × 10

FIGS. 12 and 13. Ovum C 610 removed from no. 442, January 19, 1938; attachment to the uterine epithelium had just begun. × 50. This ovum would not have been secured without the use of Gonadogen to produce ovulation in a failing animal.

CHARACTERISTICS OF THE PRIMATE EGG IMMEDIATELY PRECEDING ITS ATTACHMENT TO THE UTERINE WALL

GEORGE L. STREETER

Chairman, Division of Animal Biology
The Carnegie Institution of Washington

An attempt has been made in the following pages to point out that in primate development the first steps are accomplished by segregation of materials through the mechanism of cleavage. In this way different orders of tissues are created. The forces that appear to be responsible for their segregation are intrinsic and purely genetic rather than environmental. A second and quite different type of developmental influence is added when these primary cell groups begin to influence each other. Other types of cell groups are thus created which again in turn exert a directive influence upon each other, thereby enormously increasing the possibilities of tissue specialization. These two forces, the segregation inherent in cell division and the inductive or environmental influence of one cell group on another, overlap in varying proportions and they constitute the basic influences which must be taken into account if we are to understand the mechanism of the organization of the embryo. Moreover, these two forces are accompanied by physical cellular changes which are demonstrable under the microscope and which, if fully regarded, render the structural characteristics of the early stages of primate embryos more intelligible. The realities of this manner of analysis will certainly leave us ill content to continue leaning on the rationalizations involving two-layered and three-layered diagrammatic ground forms which have heretofore so largely dominated embryological concepts.

Introductory Remarks

It may be explained to those not familiar with the field, that one can now speak in considerable detail of the structural characteristics of early primate embryos. Though many of the lower forms have long been familiar objects of embryological investigation, detailed knowledge of primate embryology has only become available in the past few years. No one had been able to obtain sufficiently young or a sufficiently complete series of stages, nor had a suitable technique been devised to handle the material if it had been obtained. It became the problem of the Department of Embryology to surmount these difficulties. Our success in doing so has made the preparation of this paper possible.

The above success, it must be said, does not rest on any single pair of shoulders. It could only have been attained by the cooperation of several individuals. While this paper is printed under the authorship of one person it is to be remembered that in reality it is the product of a considerable group. The author plays the part of reporter and is exclusively responsible only for the interpretative matter interjected in the description of the specimens.

Before the attention of this Laboratory was diverted from human material to the embryology of closely related primates, others had already broken the ground. Descriptions had been given of early embryos of several primate forms, based on animals shot in the wild and preserved under the limitations of field conditions. The names of Hubrecht and Selenka will always be honored as pioneers in advancing this border of our knowledge and the value of such material to comparative embryology is attested in the well-known Croonian Lecture of Professor J. P. Hill. These men were of the first rank, but even in such hands it was found that primate embryological material obtained in the field had little to add regarding the structure of early stages to what was already known for the human, and it became clear that any great advance in primate embryology must come through a more complete control of the material.

Impatience with the relatively meager results being obtained with early human material and the primate material supplied by hunters led to a program which contemplated bringing into the laboratory the animals which were to supply the embryos. This was given its first impetus by Dr. G. W. Corner, who at that time was a staff member of the Department of Anatomy of the Johns Hopkins University. He showed that the rhesus monkey, *Macaca mulatta*, is a sufficiently hardy animal to be kept under observation indefinitely in laboratory cages like other laboratory animals, and this constituted the beginning of his important studies on the physiology of reproduction in monkeys. Shortly thereafter the colonization of the monkey was undertaken on a larger scale by Dr. Hartman. He, too, had an active interest in the physiology of reproduction in monkeys and having become a member of our Department entered into the problems of housing, feeding, medical care, surgical and experimental procedures, the details of which cannot be dealt with here. It is sufficient to point out that, step by step, he did succeed in converting this small, rugged, easily obtained, and relatively inexpensive primate into a laboratory animal. In some phases of the physiology of these monkeys, namely, their basal metabolism and peculiarities of body temperature, he was greatly aided by the experience and cooperation of Dr. Benedict and his associates of the Nutrition Laboratory. Through the maintenance of his monkey colony, Dr. Hartman was soon able to obtain embryos of known ovulation age at any desired degree of development from the one-celled stage on.

To be able to get normal living embryos, so to speak, on order is, however, not enough. Especially in the tiny early stages there must be someone with sufficient experience to find them and to handle them safely after they are found. This also is an essential link in the chain. This link was represented by Dr. C. H. Heuser, through whose skillful hands all specimens passed. Here, again, I must pass over the details of the technique and new devices employed by him in the fixation, photography, embedding, and sectioning, which were necessary to secure the perfection of orientation and histological excellence demanded by the importance of the material. Mention must not be omitted of the very considerable technical assistance given by Mr. O. O. Heard and Mr. C. Reather, as well as other members of the departmental staff, without which the results would have fallen short of those actually obtained.

Finally, with all this successfully accomplished there still remained the survey and interpretation of what lay under our eyes. Acknowledgment must therefore be made for the aid obtained through the observations of past and present investigators of other laboratories, and notably the new concepts of developmental processes which have arisen from experimental studies made on lower forms. The research with which this paper deals thus rests on a long list of sources ramifying within and without the Institution. Such cooperative elements in research are doubtless more prevalent than is generally realized. The present one is perhaps a fair example.

If in it an exceptionally wide stride has been made, much of the credit should go to the office of administration of the Carnegie Institution of Washington. It was from that source that the unusual facilities largely originated for the interplay of these cooperative elements.

The Pre-implantation Blastocyst

At the time we began our quest for early embryonic stages it was commonly taught that the human egg becomes implanted as a solid spherule of cells. This can be best explained by the insertion of the accompanying diagram, which is taken from Grosser, an embryologist of distinction. The diagram is designed to show that in the process of cleavage some of the cells become grouped (figure 1*a*) as an inner cell mass which is discrete from the layer of surface cells, the latter being known as trophoblast. These trophoblast cells thereon are assumed to multiply greatly, creating a thick cellular shell (figure 1*b*) which serves as the means of penetration and implantation by the ovum beneath the uterine epithelium. While this is taking place the inner cell mass becomes differentiated into an innermost central group of cells, which are the "formative" cells and will constitute the embryo proper, as against the remaining cells of the inner cell mass, which will form certain extra-embryonic structures. Subsequently the "formative" group of cells was described as subdividing into two daughter masses, each of which promptly acquires a cavity and in that manner there arise the amniotic and yolk-sac vesicle (figure 1*c* and *d*). Any medical school student giving such an explanation would be given a perfect examination rating.

In retrospect one now wonders that we could have been satisfied with this crudely mechanical interpretation. It was well known that in all other mammals studied cleavage results in a segmentation cavity that converts the solid egg mass into a "blastocyst," that is, fluid-filled vesicle. The schema (figure 1) would indicate that man, alone among mammals, is devoid of a blastocyst stage. We shall immediately see, however, that the monkey passes through a perfectly typical pre-implantation blastocyst stage and it is probable that man does likewise. It is thus clear that stage *b* of the diagram should be replaced by a blastocyst form consistent with that of other mammals. In attempting to fill in a missing stage the guess that was hazarded is far from the mark. Equally bad guesses were made regarding the yolk-sac and amnion, but those structures need not be considered at present. It is sufficient to emphasize that the early material reported in this paper opens before us an entirely new page in primate embryology and to that extent reduces the extent of developmental mechanism that has to be interpolated.

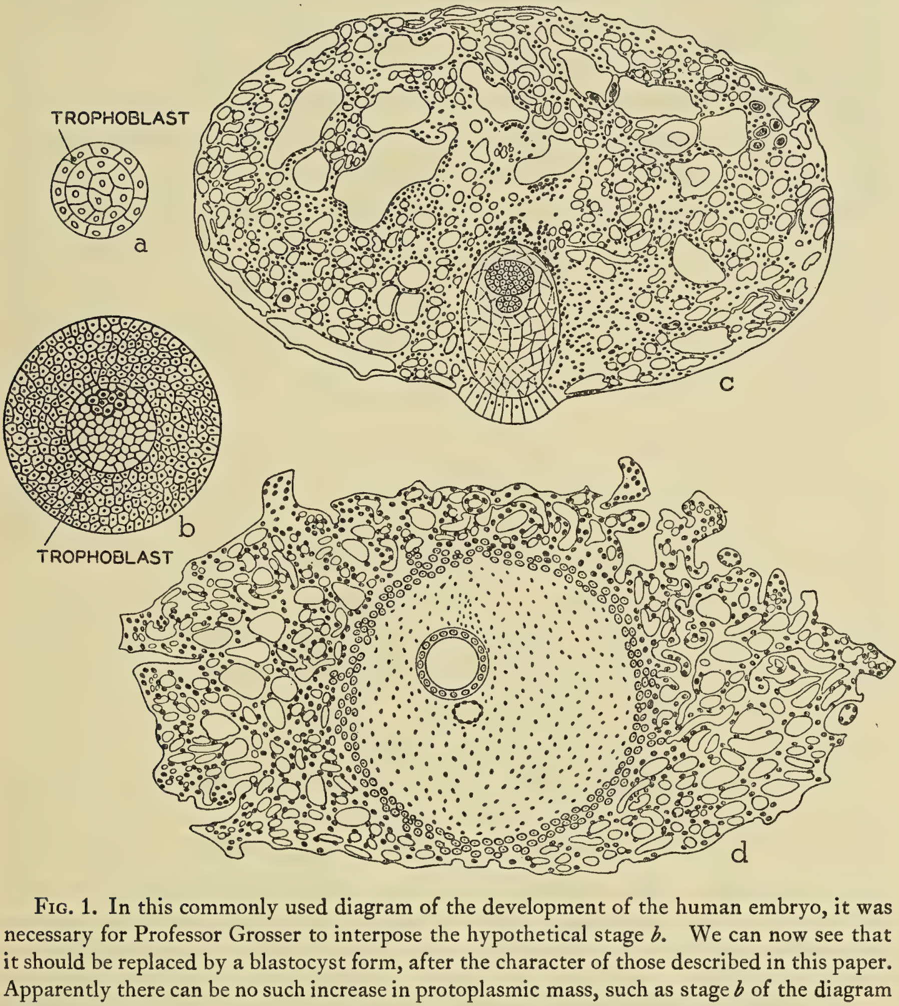


Fig. 1. In this commonly used diagram of the development of the human embryo, it was necessary for Professor Grosser to interpose the hypothetical stage *b*. We can now see that it should be replaced by a blastocyst form, after the character of those described in this paper. Apparently there can be no such increase in protoplasmic mass, such as stage *b* of the diagram calls for, until the egg is attached and begins ingestion of maternal tissue-substance.

Starting with the hitherto unknown primate blastocyst our description will deal with the egg while it is still free in the uterine cavity, and just as it is completing the degree of differentiation necessary for its attachment to and penetration of the lining membrane of the uterus. On the eighth and ninth days following ovulation such eggs can be washed out of the uterus. They are still very small and it is necessary to carry out this procedure under the binocular microscope and by means of capillary pipettes. Our description is based on four such specimens. After Bouin fixation and double embedding they have the following dimensions on the slide: no. 614, 0.14 × 0.13 mm; no. 519, 0.20 × 0.16 mm; no. 522, 0.23 × 0.22 mm; no. 613, 0.25 × 0.23 mm. Of these specimens two appear to stand out as certainly normal. One, no. 614, is definitely younger and the other, no. 522, is larger and more advanced in development. There remain two (no. 519 and no. 613) which may or may not be normal. Inasmuch as some uncertainty exists on this point, they are for the present set aside.

To aid the reader in following the description of this material, enlarged photographs are appended at the end of the paper (plates 1–3). These portray the eggs in their entirety and as cut in serial sections. On facing pages detailed descriptive matter is given which will not need to be repeated in the text. On the assumption that the reader will have glanced through these illustrations it will not be necessary to make constant reference to them.

The Eight-Day Egg

Proceeding with the younger of the two preferred specimens, no. 614, 8 days old, one finds on sectioning it that a thin homogeneous envelope, the remnant of the zona pellucida, still surrounds it. In some places signs of the disintegration of the zona can be recognized. Until it has been entirely dissolved off, it is clear that no effective contact can be made with the maternal epithelium and implantation must be correspondingly deferred. Within the zona, however, the blastocyst is actively preparing itself for that event.

Already the egg is definitely regulated as to its axis, in the sense that there is an embryonic pole of many cells and an abembryonic pole of fewer cells. In this as in its general structure it presents about the same characteristics that are found in all mammalian blastocysts. It is a very different arrangement from that represented in the hypothetical stage of the diagram (figure 1*b*).

The surface cells (trophoblast), which are to play the principal rôle in implantation, are the ones most advanced in their differentiation. They constitute the wall of the blastocyst. They, however, are not everywhere homogeneous. Those immediately at the embryonic pole behave differently from those composing the rest and larger part of the egg wall. The latter form a single layer of hexagonal cells of endothelial-like thinness and constitute a tenuous membrane enclosing the segmentation cavity. They not only serve as a container or reservoir for the blastocyst fluid but also are the agent which brings about its accumulation as is evidenced by their differentiation in structure as compared with the primitive character of the rest of the egg. It is their activity which converts the blastula into a blastocyst, and when the first lakelet of fluid appears it is proof that they are functioning; this checks with their differentiation. The subsequent increase of fluid and distention of the blastocyst is evidence of the degree of their activity.

In the region of the embryonic pole other things are happening. At least three different kinds of cells can be recognized. There are a few large cells of primitive type which appear to have undergone but few cleavages. Four of these are very large and ten others fairly large. These slowly differentiating units resemble those described by Heuser and Streeter in the pig blastocyst of a similar stage. In the pig they were found to be the ancestors of the embryo proper. It is reasonable to assume that they perform a similar function here and we may therefore provisionally designate them as "formative" cells.

In among and overlying these large primitive cells are numerous small irregularly shaped cells which are continuous with and appear to be related to their flattened trophoblastic associates which we have seen forming the larger part of the wall of the blastocyst. As has been said, these smaller irregular cells of the embryonic pole and the flattened endothelial-like cells complete the wall of the blastocyst and are thus to be grouped together as trophoblast though they serve separate functions. The further proliferation and differentiation of the polar trophoblastic cells will provide the mechanism for the attachment of the egg.

A few representatives of a third type of cell can be recognized at the embryonic pole. They spread out as a thin single-layered membrane that separates the other two kinds of polar cells from the cavity of the blastocyst. Thus at this early period the blastocyst cavity is enclosed not only by its endothelial-like trophoblast but also in the region of the embryonic pole by this third type of cell. This third type in its precocious differentiation and in its form exhibits a resemblance to the cells of the main trophoblastic wall. It is a contemporary with them and probably a close relative but is destined to serve a very different purpose. Apparently it corresponds to the primitive endoderm cell of other forms. For lack of a name that would better indicate its functional assignment it will be referred to under that well-intrenched term.

If we were to summarize the entire cell population of the blastocyst of this period we would find it varying from hour to hour in accordance with the progress of cell division, and the rate of division would vary from specimen to specimen and finally there would be a variation in rate of division of one cell type as compared with others. Such variations result in differences in the appearance of blastocysts of about the same age. It also exaggerates the differences existing between blastocysts of different mammalian species. Therefore without attaching importance to it I am giving here the approximate cell population of this particular specimen at the time it was fixed. It is as follows: (1) cavity-wall trophoblast cells, 58; (2) nodal trophoblast cells, 56; (3) primitive "formative" cells, 14; (4) primary endoderm cells, 12.

Discussion

Thus far the 8-day blastocyst like a new phenomenon has been approached in a purely objective way with pictures, measurements, and terminology and as a stage complete in itself. There remain, however, certain broader aspects and relationships which may be profitably considered. The "stage" concept is, of course, an artificial one and what we are looking at are the component parts of a blastocyst as they exist and are related to each other at the particular moment the specimen was dropped in the fixing fluid. A little earlier or a little later these relationships would be different. What we really are concerned with is the continuous story of the elements of the egg and their influence upon each other as they gradually evolve into the structures of the mature organism.

Rightly to appreciate the specimen that lies before us it must be recalled that its component elements came from one cell and in the course of cleavage of the egg they were separated out from each other. It was shown in the pig by Heuser and Streeter how in the successive cleavage stages segregation of material occurs with each cell division. This proceeds with such rapidity in the pig that in the 16-cell stage the trophoblastic material is almost completely separated from the material that is to form the embryo proper. The latter is represented by a few large primitive cells which are dividing very slowly and the former is found in small rapidly dividing cells that are already beginning to take on special form and are beginning to function in creating a lakelet of fluid or the beginning of the segmentation cavity. Thus in the pig the diverging rates of differentiation of the various elements of the egg are already striking at the 16-cell stage and since they can be seen to consist of different materials this is not surprising.

Coming back to our 8-day monkey blastocyst, as seen on plate 1, one is immediately struck by the difference in structure and rate of differentiation of its component cells. A considerable amount of segregation of materials has already been attained. We have on the one hand the large isolated blastomere-like "formative" cells which are destined to go through a number of divisions before they lose their individuality and cooperate in the formation of an embryonic disk. In contrast to these are the cells of the trophoblastic wall of the segmentation cavity. These are well along in their differentiation and are no longer independent individuals but instead have become a common tissue, a thin membrane through whose structure and function there is attained the chief characteristic of the blastocyst, namely, a vesicle distended with fluid. Such a form, with a given amount of material, offers a maximum surface for purposes of implantation.

Falling between the two extremes mentioned are the irregularly cuboidal trophoblastic cells at the embryonic pole. These are in the process of proliferation and differentiation into tissues suitable to the requirements which subsequently must be met in this region. As for the so-called primitive endoderm cells, which have already acquired the form so characteristic in all early mammalian blastocysts, it is to be noted that these cannot be derivatives of the embryonic ectoderm, for the simple reason that the embryonic ectoderm itself has not yet been organized into a unified structure. Were the order of precocity of the two cell types reversed the primitive endoderm might well give the appearance of being derived from the embryonic ectoderm. Here in the primate egg, however, we have evidence of its independent segregative origin.

Thus far we can only speak of segregation of cells in the sense of a breaking apart of the original one-cell egg. It is not until the derived daughter cells lose their individualism and become grouped into multicellular structures that we begin to see inductive phenomena and the budding off of derivative products. Simple environmental influences are doubtless already at work in the 8-day blastocyst. The increasing fluid in the segmentation cavity obviously is in part responsible for the form of the cells of the trophoblastic wall and also for the characteristics of the primitive endoderm cells. We have, however, instructive evidence in the Lewis-Gregory films of the developing rabbit egg, where fragments or herniated parts of the blastocyst wall retain much of their normal character, in spite of the altered environment. The fact that the rabbit egg can be grown *in vitro* from early cleavage stages to the completed blastocyst shows its relative independence of the uterus during that period. It may be said that what we see in the 8-day egg are primary cell types derived in the main by the processes of cleavage and segregation, and as yet one cannot profitably speak of one cell type giving off another. Had it been our aim to derive the three germ layers, we would have to admit that we have found little evidence of any such arrangement thus far. We have seen, however, that already in the 8-day egg some very fundamental things have taken place.

The Nine-Day Egg

In the pregnant uterus one may find on the ninth day either a free blastocyst or one beginning to attach. Our experience indicates that a range of 24 hours for the same degree of development is within the limits of normal variation. No especial effort has been made to determine this point. To study variations in rate of development would require a much larger number of specimens than are at our disposal and it is a problem that could be better solved with lower forms. We are limiting our attention to specimen no. 522, a blastocyst which is still free. It represents a stage next older to the one we have just been studying (no. 614).

In our selected egg of the ninth day the zona pellucida has completely disappeared; the surface of the egg can now come into direct contact with the uterine wall and the way is cleared for implantation. It will be seen that the trophoblastic cells at the embryonic pole are preparing for this event.

The chief difference between this blastocyst and the 8-day specimen is found in the cells at the embryonic pole. Those large primitive blastomere-like cells of the 8-day specimen have undergone division and with it have surrendered their individuality and now on the ninth day are grouped in a common mass. However, when compared with the overlying trophoblast it will be seen that they still retain their primitive character, and are grouped in haphazard relation to one another. There is now no difficulty in tracing in the later specimens the further conversion of this mass into the germ-disk and hence recognizing its identity as composed of "formative" cells.

Overlying the formative cell mass are the surface trophoblast cells which are actively proliferating. The area has the appearance of being crowded with nuclei two or three layers deep. The more superficial of these are the ones that are about to differentiate into a syncytial crust, the active material that will fasten to and invade the uterine epithelium. The deeper cells tend to merge with the "formative" cells and in some instances they cannot be distinguished from one another. It is from this region that the amnion cells are differentiated and they may be spoken of as amniogenic.

Ventral to the formative cell mass and clearly separate from it are the primitive endoderm cells. These constitute a definite membrane one cell thick which separates the formative cell mass from the segmentation cavity after the manner of a mesothelium lining the cavity. In this specimen the primitive endodermal membrane is not confined to the formative area but is receiving additional cells around its margins. These may be regarded either as migratory proliferations from the original primitive endoderm or as induced cells given off *in loco* by the surrounding trophoblast, and there are appearances that support the latter view. That is to say, one could either have a spreading membrane, or the same end could be at tained by a spreading wave of development *in loco*. Moreover there could be a combination of the two. But the completion of this membrane takes place subsequent to the period with which we are concerned and it need not detain us. As an aid to the reader, a three-dimensional view of this specimen is given in figure 14, plate 2, and sections through the specimen are shown in figures 15–24, plate 3, in which the various structures under discussion can be seen.

Leaving the region of the embryonic pole, we come upon a thin single-layered trophoblastic wall much like that of the 8-day egg. It forms a larger sphere and consists of more cells, which however retain their original character and continue their same function, namely, the accumulation and retention of fluid in the segmentation cavity. A note must be added here reporting the occurrence on the blastocyst wall of a localized blister-like doubling of the layers. The small vesicle (0.08 × 0.07 × 0.06 mm) thus produced lies at one side of the embryonic pole, and thus is near the fastening area. It may be a response to the implantation stimulus but since it has been seen only once its interpretation will have to be deferred.

As a summary of this specimen, a census is herewith given of its component cell population, with the same reservations, made in the younger egg, as to the significance of the exact numbers. The census is as follows: (1) cavity-wall trophoblast cells, 224; (2) polar trophoblast cells, 95 ± 10; (3) formative cells, 32; and (4) primary endoderm cells, 24.

Finally, returning to specimens nos. 613 and 519, (figures 5 and 6, plate 1) it is to be pointed out that in size and general form they correspond to the two eggs that have been described in detail. These less favored eggs both show a persisting trace of zona pellucida and they both exhibit typical differentiation of the trophoblastic shell. It is at their embryonic pole that the chief cause for doubt arises. In each of them this region appears too small, too compact, and perhaps underdeveloped. Of the two no. 613, figure 5, a 9-day egg, is possibly normal. One can clearly see primitive endoderm cells, but it is difficult to distinguish between formative cells and polar trophoblast cells, which we have seen should be recognizable at this time. It is quite possible that the appearance is due to retarded proliferation of the trophoblast cells which might have been corrected with a few hours of further development. That extent of variation might be normal. The fourth egg, no. 519, was obtained from a uterus where a development of 10¾ days was expected. At that time the ovum should be implanted and well along in development. This clinical discrepancy suggests that we may be dealing with an arrest in development and that this ovum never would have gone farther. We should expect one out of four or five eggs to be of that sort. In this case there is scant development of the embryonic pole, with poor definition of the cells. It has the appearance of a poorly preserved specimen. This, however, is not likely since it had the same treatment and fixation as the good eggs. One must conclude that the deterioration occurred in the uterus before it came into our hands. As far as can be seen in the specimen it conforms to the behavior of good eggs. Its development is too stunted to permit a distinction between formative cells and polar trophoblast cells. One can, however, distinguish a few primitive endoderm cells and this is an indication of their precocity and elemental nature. Otherwise such a specimen must be put aside as negative evidence.

Discussion

On the ninth day the blastocyst has completed its primary differentiation and has reached its maximum size. The one-cell egg has a maximum diameter inside the zona pellucida of about 0.110 mm before embedding and the nine-day blastocyst has a maximum diameter of 0.267 mm, fixed before embedding. This increase in size, however, does not mean that it has acquired more mass. In fact, during cleavage one can see that it actually becomes smaller. Up to this time then we should not speak of growth. What has happened has been a segregation of the substances of the ovum and a change in their arrangement, that is to say, a differentiation of new forms. Instead of one large cell we now have many little ones which can be grouped under at least four main types and which are so shaped and distributed as to provide a greatly increased outer surface. If these cells were now clumped together in a compact mass they would not bulk as large as the original egg. Not until the egg is implanted and can obtain the requisite substances from the uterine wall does it begin its actual growth.

The most conspicuous characteristic of the blastocyst, and the one for which it is so named, is its distention into a thin-walled vesicle. It has become a film of protoplasm stretched over a droplet of water.

This multicellular trophoblastic film constitutes the larger part of the blastocyst. Apparently this particular arrangement completes the primary purposes of the trophoblastic film within the two or three days directly preceding implantation. The fact that all mammalian eggs pass through this stage is evidence of its fundamental, though temporary, importance. As soon as the egg begins to attach to the uterine epithelium, the blastocyst wall undergoes marked alterations and assumes other functions. It is to be remembered that this part of the ovum never becomes an intrinsic part of the embryo. Instead it is an auxiliary structure that is to be classed with the devices for transport, implantation, and nourishment of the egg. The expanse of its surface just before implantation appears to be its essential characteristic and this appears to be common to blastocysts in general. There is, however, some variation among various mammals in the degree of distention and the thinness of the blastocyst wall. With further acquaintance with the eggs of many species it would seem likely that variations in wall distention could be correlated with the character of implantation. For instance, the blastocysts of the rabbit, pig, and monkey, which are not wholly or not at all interstitial in their implantation, are well distended; whereas those of the guinea pig, rat, and mouse, which are interstitial, do not become distended so much and their trophoblastic wall remains relatively thick.

The blastocyst form is not to be thought of solely in terms of the next succeeding stage in development. It is to be remembered that at all stages the embryo is a living organism, that is, it is a going concern with adequate mechanisms for its maintenance as of that time. It is conceivable that, were it possible to remove the drive to further development, these tiny blastocysts might continue to flourish as such in the uterine cavity as long as the uterus continued to supply the type of environment characteristic of this period. One should therefore regard the thin trophoblastic wall of the blastocyst as a functional membrane adapted to the immediate needs of the organism as well as being endowed with potencies for the formation of new tissues when these are called for by subsequent changes in the environment.

At the embryonic pole of the 9-day blastocyst we have something of a different order of significance. In this relatively small region there is a nodal center of active proliferation, of whose cells the more superficial ones will provide a special mechanism for invasion of the uterine epithelium and the more deeply lying ones will give origin to the embryo proper. This general region is ordinarily referred to as the inner cell mass. If such a term is to be used it should be remembered that this region does not consist of homogeneous material. We have previously seen that in this region there are at least three different orders of cells and that they are clearly recognizable as such as early as the eighth day. Thus the designation inner cell mass is a topographical term used for descriptive purposes rather than a structural one. In the late morula stage, before the first lakelets of the segmentation cavity have appeared, one can distinguish between the central mass of cells and a single layer of cuboidal cells with which a large part of the egg is covered. The latter are the trophoblast cells undergoing the differentiation that will enable them to pass fluid into the interior and thus create a segmentation cavity. The former are the undifferentiated remainder and constitute an inner cell mass from which both trophoblastic or auxiliary elements and purely formative cells are to be further segregated. Even from the outset a part of this primitive central mass extends to the surface of the egg, which means that the trophoblastic shell is not differentiated uniformly over the whole egg. In this we have an early expression of polarization of the egg and this area over which the trophoblast is not clearly differentiated as a covering layer constitutes the embryonic pole.

As the fluid in the segmentation cavity accumulates the inner cell mass can be followed as an area of retarded differentiation retaining its original position, interset in the expanding wall of the blastocyst. Here the segregation and differentiation of its various elements continues. It is in this state of development that we find it in the stage under discussion.

Conclusion

In this paper there is given a description of the primate egg as it exists free in the uterine cavity during the two days preceding implantation. It is the first time that such material has been available. Because of the importance of these four eggs the attempt has been made not only to describe them objectively but also to interpret their structure on a basis of its functional implications. It is found that much is revealed in these simple forms to give us an understanding of the subsequent processes of development, and at the same time they leave no serious gap between their own period and the late stages of cleavage when most of the mammals are close kin. An unknown period for primates has thus been covered and thereby the way has been opened to analyze the processes of development from the functional viewpoint.

Up to and during the time covered by this period, it has been shown that development consists almost exclusively of the segregation and differentiation of the primary tissues. Environmental influences of one tissue upon another do not begin until about this time. With the superimposition of environmental factors the possibilities of new types of tissue become enormously increased and the embryological problem correspondingly complicated. However, with the earlier events thoroughly understood the organization of the embryo and its auxiliary membranes can be analyzed with greater accuracy. For that reason these eggs will serve as an important key to primate development.

Embryos at every stage, even as far back as the unattached blastocyst, are engaged in two lines of activity. In the first place it is obvious that the embryo must undergo cell multiplication, tissue segregation, and in one way and another the differentiation of complicated organs, characterizing growth and development. But underlying all this there is the requirement that, in their rôle of living organisms, it is necessary that embryos maintain themselves at their respective biological level and that their structure be so designed for each developmental stage as to insure an adequate physiological performance. It would follow that their design would be such that they could continue living indefinitely at any respective stage, so long as no change in themselves or their environment rendered it inadequate. In short the embryo not only develops but also lives. It becomes the function of the embryologist, therefore, in his study of the objects revealed to him under the microscope, to distinguish the manner in which these two requirements are met. Among other things he will see that as the requirements for existence progressively change there are developed a series of temporary devices with which those particular needs are met. When the requirement is past the device is altered or permanently discarded. Such ephemeral structures are already present in the blastocyst, the interesting stage with which this paper has been concerned.

BIBLIOGRAPHY

Grosser, O. 1927. Frühentwicklung. Eihautbildung und Placentation. Bergmann, München.

Heuser, C. H., and G. L. Streeter. 1929. Early stages in the development of pig embryos, from the period of initial cleavage to the time of the appearance of limb-buds. Carnegie Inst. Wash. Pub. No. 394, Contrib. to Embryol. vol. 20, pp. 1–30.

Hill, J. P. 1932. Croonian Lecture. The developmental history of the Primates. Phil. Trans. Roy. Soc. London, s. B, vol. 221, pp. 45–178.

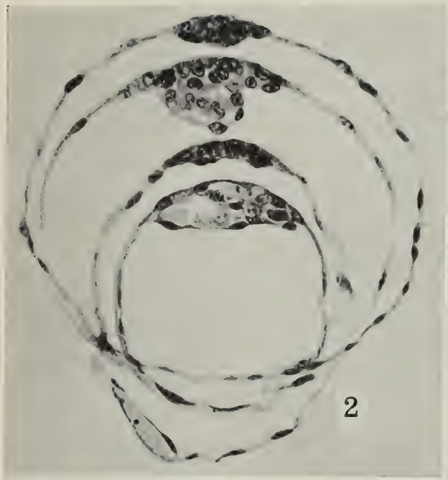
PLATE 1

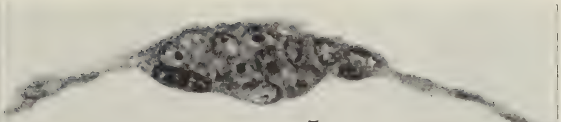
FIG. 2. Sections of four macaque ova, at the eighth and ninth days, when they are still free-blastocysts. Shortly after this they will invade the tissues of the uterine wall. In order from above downward, they are from the following specimens: C. 613; C. 522; C. 519; C. 614. They are all enlarged × 200.

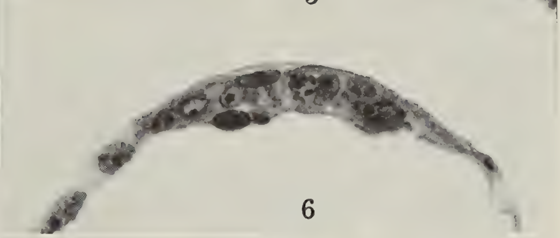
FIG. 3. Surface view of the eight-day macaque ovum (C. 614) showing character of cells constituting its trophoblastic wall. Enlarged × 250.

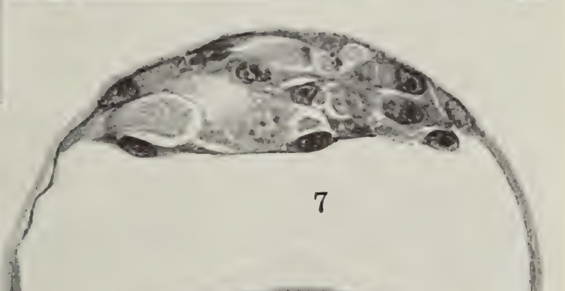
FIGS. 4 AND 7–10. Five successive sections through the embryonic pole of the same ovum shown in fig. 3, and all enlarged × 500. These photographs show that, due to the process of segregation and differentiation, four kinds of cells are already discernible by the eighth day, namely 1, trophoblast cells of the blastocyst wall (fig. 4); 2, trophoblast cells of the embryonic pole; 3, large primitive "formative" cells (figs. 4 and 9) which, when they have divided a few more times, will unite in the formation of the embryonic ectoderm; and 4, primary endoderm cells, forming a membrane beneath other cells of embryonic pole (fig. 7). In fig. 4 it can be seen that the zona pellucida is still present as a homogeneous membrane surrounding the eggs, though beginning to disintegrate.

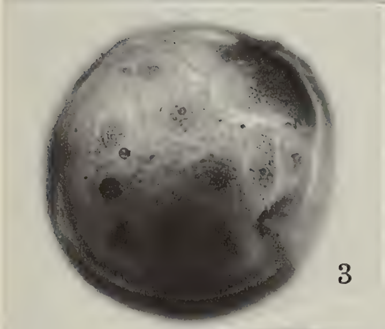
FIGS. 5 AND 6. Sections through the embryonic pole of embryos C. 613 and C. 519 respectively. Both are enlarged × 500. The trophoblastic walls of both of these specimens appear normal and both of them have normal appearing primary endoderm cells. The other cells of the embryonic pole, however, appear deficient in cytoplasm to an extent which suggests that they are abnormal.

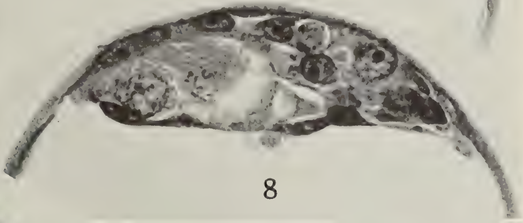

2

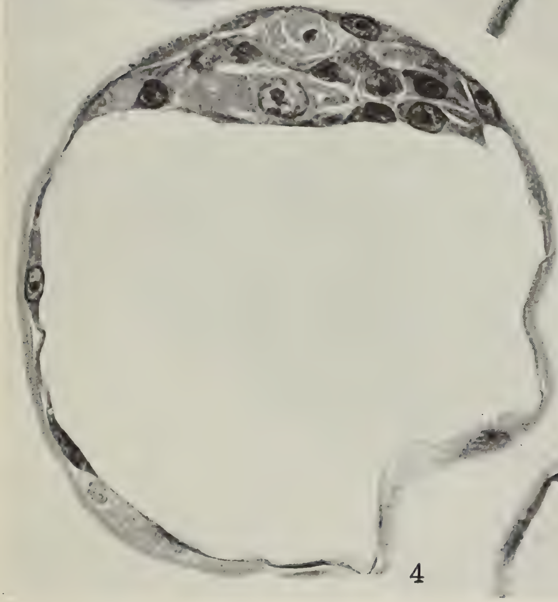

5

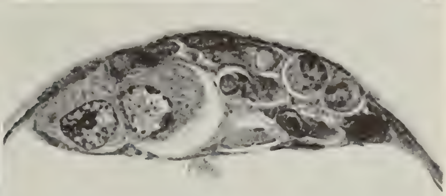

6


7


3


8


4


9

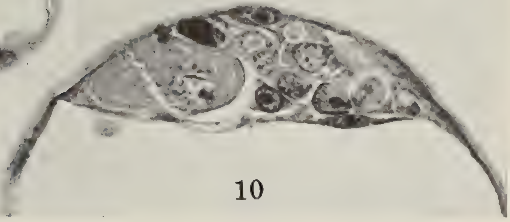

10

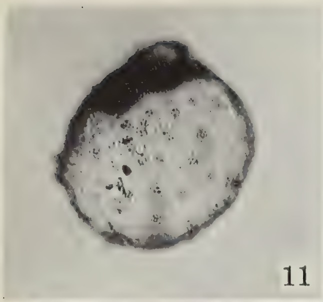
PLATE 2

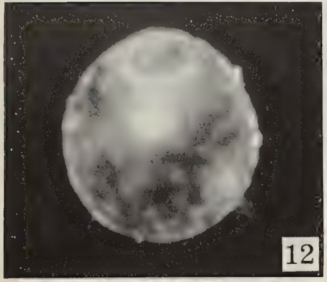
All figures on this plate show the same 9-day ovum, C. 522.

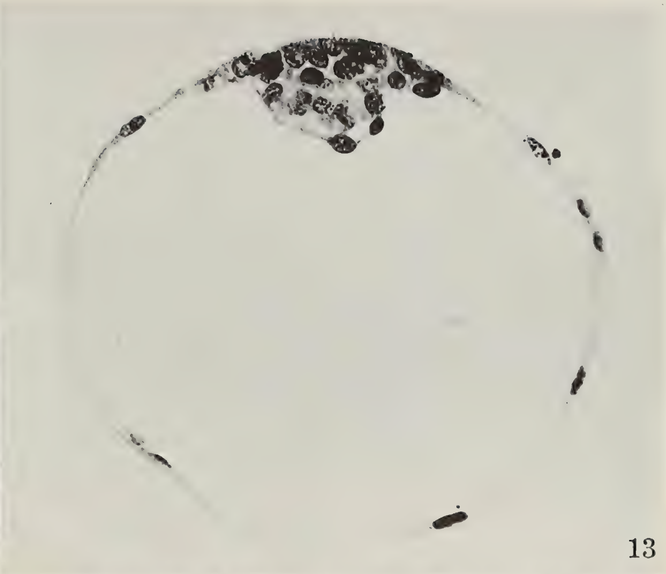
FIGS. 11 AND 12 show intact ovum with transmitted and direct light, enlarged × 100.

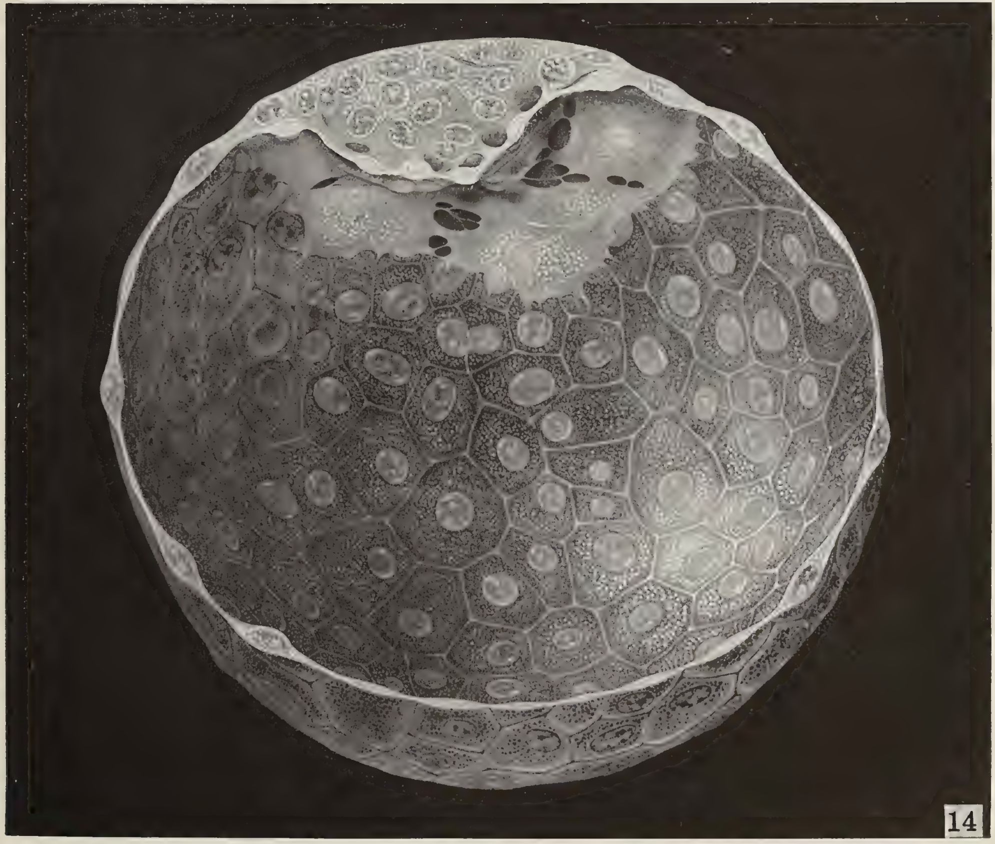
FIG. 13. Section of ovum, showing the four groups of cells already noted in the 8-day egg. The chief difference is the further division of the "formative cells" and their assembly in an embryonic ectodermal mass. Enlargement × 300.

FIG. 14. Reconstruction of one-half of the egg, enlarged × 400. Though the drawing is in part interpretative, the number and position of the nuclei were accurately reconstructed from serial photographs. Of special interest are the primary endoderm cells that form a membrane separating the other cells of the embryonic pole from the cavity of the blastocyst.


11


12

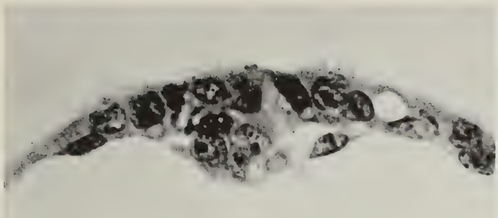

13

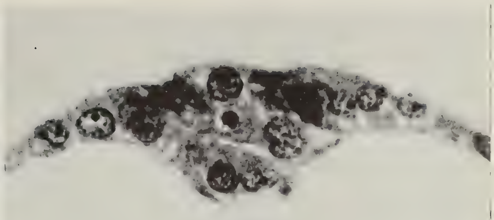

14

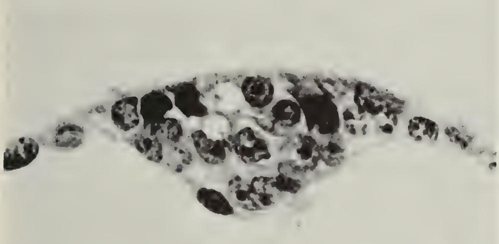
J. F. Didusch, Del.

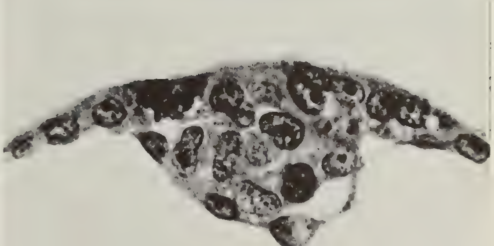
PLATE 3

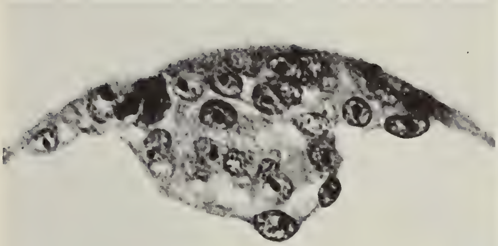
Sections of the 9-day ovum shown on Plate 2, C. 522, all enlarged × 500. These are successive sections and form a complete series through the main part of the embryonic pole. The cells destined to form the embryo can be recognized as a central mass in figs. 18–20. On each side of this region the sections become tangential to this mass and its contours are less discrete. We thus have in the 9-day ovum a distinction between the embryo-forming cells and all the others which go to form various auxiliary parts of the ovum, but which never become a definitive part of the embryo. The auxiliary cells are relatively precocious as it is necessary for them to provide the mechanisms of attachment and of nourishment of the contained embryo.

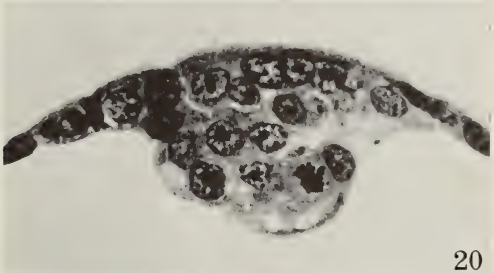

15

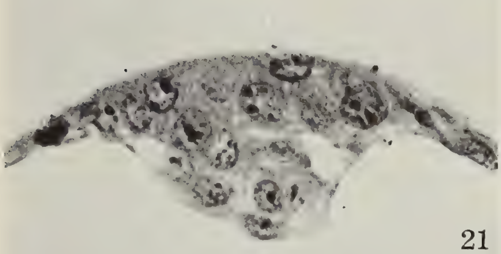

16

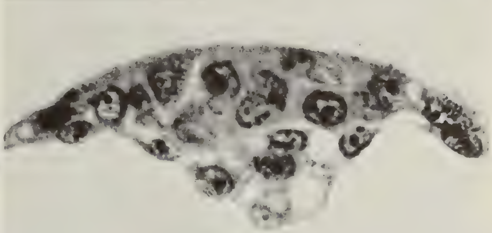

17

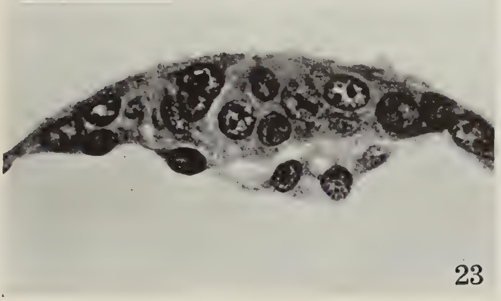

18

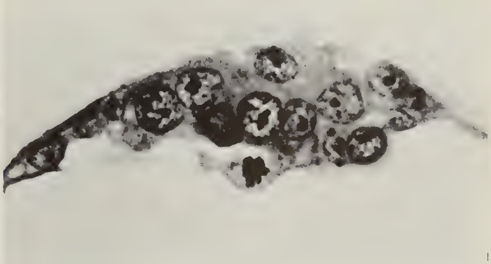

19


20


21


22


23


24

THE INFLUENCE OF NUTRITION UPON THE CHEMICAL COMPOSITION OF THE NORMAL BODY

HENRY C. SHERMAN

Mitchill Professor of Chemistry, Columbia University
Research Associate, The Carnegie Institution of Washington

A useful characterization of what food does in nourishing the body is: (1) that it yields energy for the body's activities; (2) that it provides the materials for the construction and upkeep of tissues; and (3) that it supplies the substances through which, or through whose derivatives, the body's self-regulatory processes are maintained.

It was the energy aspect which first inspired the interest and support of the Carnegie Institution, and to which the researches of its own Nutrition Laboratory have thus far been primarily directed. Here the methods and viewpoints of chemistry, physics, and physiology are merged in the study of the nutritional needs of the body as an energy transformer. Naturally we look to Dr. Benedict to discuss both the experiments carried on within the laboratory of which he has been for three decades the director, and the clinical and field work effected through his cooperation with individuals and organizations, intensively in his own community and extensively throughout the world.

While thus supporting a unique research laboratory of its own for the study of the energy aspect of nutrition, the Institution has also cooperatively fostered researches in other aspects of this science.

It aided Dr. Osborne's investigations in the isolation and descriptive chemistry of the proteins at the Connecticut Agricultural Experiment Station in New Haven; and then supported the nutritional studies of these proteins, which he conducted jointly with Dr. Mendel of Yale. This work of Osborne and Mendel has played an outstanding part in the establishment of the principle that the nutritive value of a protein is essentially dependent upon its amino-acid makeup. It also showed clearly and strikingly that some of these amino acids are individually indispensable and are used in fixed proportion in the construction of the growing body. Lysine, for instance, when furnished by the diet in restricted proportion, thereby becomes the growth-limiting factor—the body growing only so fast and so far as is consistent with the normal composition of its kind, in this respect.

Thus the main bulk of the structural and replacement material of body tissue was shown to come from the proteins of the food; and many of the individual chemical factors within protein molecules which determine their nutritional efficiencies were identified and quantitatively investigated.

Meanwhile, the source of the energy transformed in the processes of nutrition had been conclusively traced to, and quantitatively accounted for in, the organic foodstuffs which undergo cleavage and oxidation in the body.

But there remained the question, how to account for the fact that these chemical reactions of cleavage and oxidation go on so much more rapidly in the body than at the same temperature and the same concentrations of reactants *in vitro*. The word catalysis applies, but does not in itself supply any explicit information: what are these catalysts? Prominent among them are the hydrolytic enzymes which digest the foodstuffs, and those which function correspondingly in plant physiology, as illustrated by references to the amylases in several of Dr. Spoehr's reports. In the study of the chemical natures and nutritional antecedents of enzymes of this type the Institution cooperated with Columbia University in an investigation, the findings of which, together with those of simultaneous researches on other aspects of the problem, notably Northrop's at the Rockefeller Institute, showed these to be of essentially protein nature. Thus in their material makeup and antecedents these digestive enzymes, as well as such other nutritional catalysts as insulin, thyroxine, epinephrine, and glutathione, are essentially derivatives of the digestion products of the food proteins.

Thus nutrients which are prominently concerned in the building and upkeep of body tissue are found also to function in the energy, and in the regulatory, aspects of nutrition. Equally significant are the interrelations of the structural, the oxidative or energy-transforming, and the regulatory functions of the mineral elements and vitamins concerned in the body's nutritional processes.

Plainly the threefold characterization cited at the beginning of this paper means not three separate processes but rather three aspects of a unified nutritional process, which in turn is one general aspect of the life process as an interrelated whole.

Here as elsewhere, science is constantly striving to make itself more and more exact; and is frequently finding that the qualitative and quantitative phases of scientific progress interlock. The obvious sequence of first forming a qualitative acquaintance with the factors concerned in a complex process, and then proceeding to measure quantitatively the influence of each, often proves in actual research to constitute one cycle of a spiral progress. For the quantitative phase of a far-reaching scientific development is often so much more than merely measuring, that only through its precision can science gain the deeper insights which reveal the unforeseen; so that qualitatively new concepts arise from quantitative research.

This is the present position of that whole aspect of nutritional research which centers, in functional terms, upon the body's regulatory processes and internal environment; or in terms of the nutrient intake, chiefly upon the mineral elements and vitamins.

Fully aware that it is not yet certain just how many mineral elements and vitamins are involved in the nutritional process, we nevertheless have a sufficient acquaintance with the qualitative aspects of the problem to enable us to go on to that type of research in this field which is more rigorously precise and quantitative. Thus to the qualitative question, What things are involved? we add, with reference to such of these things as show greatest probability of being "limiting factors," the quantitative questions, How much do different foods supply? How much does the nutritional process require? and more ambitiously, How much gives the best results?

Here nutritional research not only seeks to understand the normal processes and to restore abnormalities to normal; but considers also the more constructive question, whether by finding (and thereafter following) the optimal levels of intake of those nutritional factors which prove most influential in this connection, it may be possible to effect significant improvement in the life process by the more scientific guidance of the relative proportions in which active chemical factors are introduced into the body through its food.

Through the cooperation of the Carnegie Corporation of New York and the Carnegie Institution of Washington it has been possible during the past few years to develop in the Department of Chemistry at Columbia University a more comprehensive and rigorously quantitative plan of nutritional research of this type. Accurately controlled feeding experiments upon large numbers of laboratory animals of known antecedents have been continued throughout entire life cycles and also through successive generations. An Institution lecture of March 1936 summarized the more general findings up to that time and suggested very briefly a few of their broader human implications.

What follows, therefore, deals more specifically with the work of the past two years.

It may now be regarded as conclusively established, both by calcium-balance experiments with children and by analyses of the bodies of laboratory animals, that the level of calcium intake materially influences the calcium retention and resulting calcium content of the growing body. Thus with respect to calcium the body does not "obey" what is sometimes called "Liebig's law" of constancy of composition, as it did with respect to lysine in the experiments of Osborne and Mendel mentioned above. The general principle that the body tends to grow to a certain normal chemical composition for its kind evidently applies more strictly to some of the essential factors of nutrition than to others; or, if one prefers, the fact may be expressed by saying that for some of the essential chemical constituents of the body the quantitative variations consistent with normality are much wider than for some other constituents. For the present each constitutes its own problem; though perhaps not quite literally an entirely separate problem, for some interrelationships may be established as our knowledge becomes more complete.

Our experiments in this field have thus far dealt primarily with calcium.

By continuing the controlled comparative feeding of the experimental animals into the second generation (or longer) and analyzing at definite ages the offspring from the families of the same hereditary background and the same environments except for differences in the mineral contents of their food, one obtains more conclusive evidence than has hitherto been available upon the far-reaching biochemical problem of the extent to which nutrition influences the composition of the body among the normal members of a species.

The present investigation deals primarily with the constructive aspect of this problem. It starts with a dietary or food supply which is adequate as shown by the fact that families of experimental animals, of the same strain used in the work here reported, are thriving in the forty-fourth generation on this diet in the hands of Dr. H. L. Campbell of our laboratory. This basal dietary (the Diet 16 of our laboratory records and the Diet A of some of our previous papers) is however of only slightly more than minimal adequacy, for permanently successful nutrition generation after generation, in its calcium content, which is 0.20 per cent of the dry food mixture. In contrast with this low-calcium adequate dietary, we have fed dietaries like it in other respects but containing about threefold and fourfold more calcium, viz. 0.64 per cent (Diet 168) and 0.80 per cent (Diet 169), respectively.

The offspring are found to differ strikingly in their body calcium content during a relatively long period of rapid growth. Rats one month old from the families on liberal-calcium dietaries have a percentage of body calcium such as is attained only at five to six months by their cousins from families whose food contains only a minimal-adequate amount of calcium. Through the period of most rapid growth the former have about one-half more body calcium than the latter. Thereafter the difference diminishes as all approach the adult condition of stabilized calcium content, but even as adults those from the families of liberal-calcium food supply still contain both larger amounts and larger percentages of body calcium than do those from families whose food is of only slightly more than minimal-adequate calcium content.

Thus from a starting point already within the normal zone, enrichment of the dietary in its calcium content is shown to have an effect upon the calcium content of the body which, while not apparent in the offspring at birth, is very pronounced during the period of most rapid growth and development, and remains significantly measurable in fully adult life.

Such differences in the chemical composition of the normal body due to differences of dietary intake no larger than occur commonly both in the experiences of animals in nature and in those of people in civilized life, can hardly fail to have far-reaching implications. To what extent, for instance, the nutritionally induced differences in the body's calcium stores will bear upon the sex differences in blood calcium levels which Dr. Riddle and his coworkers are studying, is an interesting problem for the future.

The most striking feature of our results thus far is the unexpectedly large and long effect of the calcium content of the family food supply upon the percentage of calcium in the bodies of the growing young; greatest at the period at which growth is most rapid, and significant throughout the whole period from the end of infancy to the attainment of adulthood.

Naturally the question arises whether the slower or the more rapid attainment of the normal adult calcium percentage is more advantageous to the well-being of the body considered from the viewpoint of the entire life cycle, of the whole life history.

Animals of nutritional histories parallel to those analyzed in the experiments above mentioned are therefore being continued on their respective diets throughout their natural lives. This part of the work is still in progress, but the results thus far leave no room for doubt that the more liberal levels of calcium intake will prove more favorable to adult vitality or fitness and to length of life.

Thus the greater retention and higher percentage of calcium in the bodies of the young animals whose own and whose parents' food had been of liberal calcium content appears clearly to be an entirely advantageous expediting of the normal calcification process, beneficial both in its immediate and in its ultimate results.

In view of the important differences in body composition and life history thus induced by calcium enrichment of our already adequate but sub-optimal Diet A, we are now studying the effects of increasing the calcium intake when the original diet is more nearly optimal in other respects. For while it is probable that nutritional improvements will usually be additive, there is also a well-recognized hypothesis that calcium, besides performing its individual nutritional functions, may also play a regulatory or stabilizing rôle in cases of imbalance among other factors. The calcium intake needed to bring the body to its plateau of optimal performance might therefore conceivably differ according to the relative proportions of other things in the diet.

We hope also to extend our study of the influence of nutrition upon the chemical composition of the body to other factors, especially to vitamin A and riboflavin, both of which appear, like calcium, to be increasingly beneficial at successively increased levels of intake up to probably three or four times the level of minimal adequacy. And the effects of liberal intakes of each of these three factors upon the concentration levels of certain other factors in the body should likewise be studied. We already have indications that the calcium intake may influence the body's retention of iron; and there are suggestions worthy of further study that the riboflavin intake may influence the hemoglobin content of the body at a given age. There should be further experimentation on each of these points and on the connection between them, as well as on other nutritional factors and their possible interrelationships. It may or may not become necessary to a satisfactory completion of this line of research to experiment also upon the effects of different intakes of such electrolytes as sodium chloride or soluble phosphate upon the concentrations of the corresponding elements or ions in the body.

Fortunately the three nutritional factors which have appeared most promising to study in this connection because they present such wide margins of beneficial surplus above actual need, or such large areas for investigation between the minimal and optimal levels of intake for normal nutrition, are also representative of three widely different types of solubility. Calcium is chiefly held in the body in a chemical form of sparing solubility, but largely in a physical form, especially in the trabeculæ, which tends to make calcium regulation more perfect the higher the calcium content of the body. Riboflavin, on the other hand, is freely soluble in water; and vitamin A is freely soluble in fat. When these three substances have been thoroughly investigated as to the influence of their levels of nutritional intake upon the concentrations in which they and other substances are held in the body, we shall doubtless be able to formulate more fully the scientific explanation of the unexpectedly constructive advances in well-being which have been induced by the more scientific adjustment of an already adequate dietary.

It need hardly be said that such gentle chemical treatment as a shift of quantitative proportions of normal nutrients, and all within the range of everyday experience, is not expected to change the germ plasm. When, therefore, we find that such nutritional improvement of the already normal shows its effects more fully in the second generation than in the first, the explanation is probably to be sought in part through the simple conveyance of a better-proportioned nutritional intake to the offspring from an earlier starting point, and in part through the concept of the influence of nutritional intake upon internal environment. If a more scientifically balanced dietary has improved the mother's internal chemistry, even in the sense of keeping its normal fluctuations more constantly in the more nearly optimal ranges, this is doubtless beneficial to her in her own life history anyway, and during pregnancy is beneficial to her unborn offspring in that it is developing in a superior environment.

The recent development of much more comprehensive and precisely controlled comparative feeding methods than have hitherto been available in nutritional research has afforded a new insight into the constructive possibilities of improved quantitative balance among the nutrient factors of normal dietaries.

By way of scientific explanation we now have evidence that the quantitative chemical composition of the normal body is subject to a wider range of modification by nutritional intake than has hitherto been supposed. Liebig's rightly influential doctrines need now no longer retard the nutritional improvement of a body composition which is already within the normal range. That Claude Bernard's postulate of *fixité* of the organism's internal environment must also be held less rigidly than in the past is indicated both by this same evidence of the influence of nutrition upon body composition and by physico-chemical principles according to which one logically expects that differences in the amounts and proportions of active factors introduced into any system will influence some of its concentration levels and dynamic-equilibrium points.

Hitherto, through our undiscriminating acceptance of the postulated fixity of the internal environment of the normal organism, we have been delayed in cultivating the field of nutritional research which the internal environment offers. Our formal definition of environment has been broad enough to include all other than hereditary factors; but our actual thinking about environment has dealt with surroundings. Now that we realize that within the normal range, not only remedially but also constructively, nutritional intakes do influence the chemical factors of the internal environment, a fairly new and very promising field of research is opened. For this nutritionally modifiable internal environment is of course the whole chemical aspect of that *milieu* which directly environs, and thus immediately conditions, the actual processes of life.

THE EFFECTS OF VOLUNTARY CHANGES IN BREATHING ON HUMAN RESPIRATORY EXCHANGE

THORNE M. CARPENTER

Acting Director, Nutrition Laboratory
The Carnegie Institution of Washington

ALTHOUGH there have been many investigations on hyperventilation, in most instances the hyperventilation has been of such a character that it is not comparable with the slight changes that may be produced in breathing, due either to drowsiness or to small amounts of activity. It is generally believed that, when the pulmonary ventilation is greater than called for by the metabolism going on at the time, there is an extra elimination of carbon dioxide which results in an abnormal respiratory quotient, i.e., one that is too high. If compensations for these slight changes in ventilation take place rapidly, such changes would be of no consequence in the determination of the respiratory exchange, particularly the respiratory quotient, in periods of 15 minutes or longer. However, if the organism does not return promptly to equilibrium, it is conceivable that such changes may produce not only abnormally high respiratory quotients in the period in which they occur, but also abnormally low respiratory quotients during the compensation periods following. The relations of overventilation to the respiratory quotient, the oxygen consumption, and the carbon dioxide elimination have been experimentally studied by Yoshinaga,[1] Sylla,[2] and Sturgis.[3] Space does not permit, however, a review of the many studies on hyperventilation.

The following investigation was made as part of a program of the Nutrition Laboratory to study the factors affecting the respiratory quotients of man and animals. The respiratory quotient is an index of the character of the material burned in the body. In order to interpret correctly the determinations of the respiratory quotient, it is important to have knowledge regarding any factors that may affect the quotient. The effects of small voluntary changes in breathing on the respiratory exchange were studied on 5 days with a trained subject, J. C., in a post-absorptive condition. On each day there was, first, a series of normal periods, followed by a single period in which the subject was instructed either to breathe more deeply than usual or to hold his breath voluntarily at intervals in order to produce either excessive ventilation or underventilation. Immediately thereafter observations were made during periods in which the subject voluntarily followed his own desire with regard to breathing, and thus a compensation could take place. The respiratory exchange was measured by the open-circuit method of Carpenter and Fox,[4] and periodically (at 5- to 7-minute intervals) determinations of alveolar carbon dioxide were also made by the method of Carpenter and Lee.[5] In the last two experiments of the series, the expired air during the abnormal and first recovery period was collected in gasometers.

---

[1] K. Yoshinaga, *Fukuoka-Ikwadaigaku-Zasshi*, vol. 27, p. 101, 1934.

[2] A. Sylla, *Med. Klin.*, vol. 31, p. 314, 1935.

[3] C. C. Sturgis, unpublished work, cited by G. Lusk, in "Science of Nutrition," 4th ed. pp. 96–97. Philadelphia and London, 1928.

Examples of the results of observations of this character are shown in table 1. On March 25 the preliminary normal periods consisted of two of 15 minutes each. These were followed by a period of 5 minutes in which the subject made an effort to overventilate. At the end of the 5-minute period he was allowed to breathe at will during two 15-minute periods. Then there was a period of 5 minutes' duration in which the subject voluntarily restricted his breathing so as to produce an underventilation. Two recovery periods followed, and again there was a 5-minute period of overventilation and later another period of shallow ventilation. On April 3 there was a normal period, followed by a period of deep breathing with somewhat larger ventilation than took place on March 25.

The overbreathing and the underbreathing had noticeable effects on the carbon dioxide elimination. In the first overbreathing period on March 25 there was a marked rise above normal of nearly 100 cc per minute. On the contrary, in the first period of shallow breathing there was a slight fall below normal in the carbon dioxide.

---

[4] T. M. Carpenter and E. L. Fox, *Arbeitsphysiologie*, vol. 4, p. 527, 1931.

[5] T. M. Carpenter and R. C. Lee, *Jour. Nutrition*, vol. 6, p. 37, 1933.

In the last period of shallow breathing on March 25 there was an even greater fall. Furthermore, not only was there a change in the period of voluntary abnormal ventilation itself, but also in all periods immediately following the abnormal period there were changes in the carbon dioxide elimination. These were in the direction of compensation. For example, the first period of overbreathing on March 25 was followed by a lowering of carbon dioxide from the normal level of 181 to 154 cc, and the second shallow breathing period was followed by a rise in carbon dioxide from the initial normal level of 181 to 194 cc. On April 3 the recovery took three 10-minute periods, as in none of these recovery periods was the carbon dioxide elimination at the level preceding the overventilation period. Hence there is a marked time element in the recovery process after slight overventilation.

TABLE 1. THE EFFECT OF VOLUNTARY CHANGES IN DEPTH OF BREATHING ON THE RESPIRATORY EXCHANGE OF J. C.
(Values per minute)

Kind of period	Duration	Carbon dioxide	Oxygen	Respiratory quotient	Respiration rate	Ventilation	Average alveolar carbon dioxide
	minutes	*cc*	*cc*			*liters*	*per cent*
March 25, 1935:							
Normal	30	181	228	0.79	16.2	6.82	6.14
Deep breathing	*5*	*276*	*256*	*1.08*	*18.4*	*10.57*	*5.24*
Recovery	15	154	214	0.72	16.7	6.21	5.46
Normal	15	176	224	0.79	16.7	6.80	5.94
Shallow breathing	*5*	*164*	*210*	*0.78*	*15.4*	*5.83*	*5.85*
Recovery	15	182	228	0.80	15.9	6.80	6.02
Normal	15	178	221	0.80	16.3	6.89	6.08
Deep breathing	*5*	*232*	*239*	*0.97*	*18.0*	*8.81*	*5.31*
Recovery	15	163	223	0.73	17.3	6.18	5.66
Normal	15	170	217	0.78	16.3	6.68	6.19
Shallow breathing	*5*	*139*	*190*	*0.73*	*12.0*	*5.13*	*6.68*
Recovery	15	194	235	0.82	16.7	7.24	6.38
Normal	15	180	226	0.80	15.9	6.98	5.96
April 3, 1935:							
Normal	15	179	230	0.78	16.2	6.87	6.09
Deep breathing	*5*	*330*	*265*	*1.24*	*20.8*	*14.34*	*4.89*
Recovery	10	127	199	0.64	16.4	5.16	5.24
Recovery	10	163	220	0.74	16.9	6.43	5.97
Recovery	10	171	222	0.77	16.2	6.51	6.02
Normal	15	183	233	0.78	15.9	6.99	6.00

The oxygen absorption also, surprisingly, underwent alterations as the result of voluntary changes in breathing. In the first deep breathing period on March 25 there was a rise above normal of 28 cc per minute. There was subsequently a slight fall. When there was shallow breathing there was a lower oxygen consumption than normal, as shown in the first shallow breathing period of March 25. The other two periods of deep breathing and shallow breathing gave similar results. On April 3, 30 minutes, at least, were required before complete recovery in respect to oxygen absorption was attained. Hence not only was the carbon dioxide elimination affected by voluntary changes in breathing, but also the oxygen absorption. This has not always been recognized in discussions of the effects of voluntary changes of breathing on the respiratory exchange. Usually only the carbon dioxide elimination is discussed.

These voluntary changes in breathing produced noticeable alterations in the respiratory quotient in the periods of abnormal breathing and also in the periods of recovery following those of abnormal breathing. For example, the first overventilation period on March 25 resulted in a respiratory quotient of 1.08. This was higher than might occur in normal post-absorptive experiments, but not impossible. It was followed by a compensatory lower quotient of 0.72. That the subject had recovered from the overventilation was shown in the normal period following the recovery period, in which the respiratory quotient was 0.79 as compared with the preliminary normal quotient of 0.79. The first period of shallow breathing did not result in a very marked change in the respiratory quotient, nor was there a compensatory change in the first recovery period after the shallow breathing. The second period of deep breathing resulted in an elevated quotient, which might readily be found with subjects who are untrained. It was followed by a compensatory lower quotient of 0.73. The second period of shallow breathing resulted in a low quotient of 0.73, followed by an elevated quotient of 0.82 as compared with the normal value of 0.79. On April 3 there were marked changes in respiratory quotient, three 10-minute periods following deep breathing having been required before the normal respiratory quotient was regained.

The respiration rate was not seriously affected by these voluntary changes in breathing. When deep breathing was carried out, the subject made an effort to increase the depth rather than the rate of breathing, and for the most part this was accomplished, although all the overventilation periods showed slight increases in respiration rate. In the recovery periods, however, there were but small changes in the rate, and consequently the recovery was carried out by changes in depth of each respiration.

That there were real changes in total ventilation is shown by the volume per minute of pulmonary ventilation. In the first deep breathing period on March 25, this rose from 6.82 to 10.57 liters per minute, and in the recovery period following there was a decided drop in the ventilation per minute. All the other abnormal and recovery periods on March 25 showed likewise decided changes in ventilation, with the exception of the recovery period after the first period of shallow breathing. On April 3 the breathing during the excessive ventilation was more than doubled in volume. However, the volume of ventilation in the following compensatory period of 10 minutes, although lower, was not so low as one might expect from the excessive ventilation during the 5-minute period. This again points to the fact that the process of recovery is a slow one and not accomplished immediately. If it were accomplished immediately, the volume of ventilation of the lungs in the first 10-minute recovery period would have been markedly lower than it was on April 3. The determination of the total lung ventilation is of aid in interpreting the changes in respiratory quotient. This is one of the advantages of respiration apparatus in which provision is made for this measurement.

Davies, Haldane, and Kennaway[6] and many others have shown that voluntary hyperventilation is always accompanied by a marked decrease in the alveolar carbon dioxide. The last column of table 1 gives the average percentage of alveolar carbon dioxide for each period. Deep breathing produced a marked fall in the alveolar carbon dioxide in all three such periods. The greatest fall during deep breathing was, on the average, to 4.89 per cent on April 3. Also in the recovery periods following the deep breathing periods the alveolar carbon dioxide was lower than in the initial normal periods. The lowest single value (3.59 per cent) occurred in the first minute of the recovery period following the period of deep breathing on April 3. On the contrary, when there was shallow breathing, the average alveolar carbon dioxide was elevated in only the second shallow breathing period and its following recovery period. The maximum percentage in a single sample was 7.18 at the first minute of the last recovery period on March 25. It was much harder for the subject to hold his breath or breathe at a shallow depth than it was for him to overventilate voluntarily. The changes in alveolar carbon dioxide were in the expected direction and show that determinations of the alveolar carbon dioxide would be of aid in interpreting the respiratory quotient.

---

[6] H. W. Davies, J. B. S. Haldane, and E. L. Kennaway, *Jour. Physiol.*, vol. 54, p. 32, 1920–1921.

The results obtained in other experiments, the details of which are not given, were similar to the examples shown in table 1.

If these effects on the carbon dioxide elimination and oxygen absorption are due to a disturbance of gaseous equilibrium in the body and a subsequent recovery, it is evident that if the experiment were long enough, the extra carbon dioxide eliminated and the extra oxygen absorbed in the overventilation period ought to be compensated for by the decreases in these factors that occur in the periods following. Calculations have been made of the balances between the extra elimination of carbon dioxide and extra absorption of oxygen in the overventilation periods and the deficit in these two values in the recovery periods following overventilation. Similar calculations have been made for the periods of shallow breathing. The problem of how this calculation should be made was a difficult one. The method finally adopted was based on the assumption that the normal base line of metabolism did not change throughout the day's experiment and that this normal base line would therefore be indicated for each of the measured factors of the respiratory exchange by summing all the values obtained for that factor in all the periods for the entire duration of the experiment, and calculating from this sum the average value per minute. The average values thus obtained have served for calculation of the excess carbon dioxide elimination and the excess oxygen absorption in the periods of deep breathing, and for calculation of the differences between the expected normal values and those found in the recovery periods after the periods of overventilation. The same procedure was used for the periods of shallow breathing and subsequent recovery. The results are given in table 2.

There was a nearly exact compensation and balance in the carbon dioxide elimination. This was particularly true in the last period of shallow breathing on March 25 and the two periods of excessive breathing on April 3. In the calculation of these balances it was assumed that the normal base line of the metabolism was in no way variable during the whole series of periods. This, however, is not probable, and some of the slight differences in the balances must be due to the actual slight differences in the base line from period to period. It is well known that if respiratory exchange measurements are made on a subject in successive periods, the results are not exactly constant from period to period. It is only with an exceedingly well-trained subject having marked self-control that such constant values can be obtained. Therefore, it is not surprising that in some cases the balances in these experiments were not perfect. The actual differences per minute between the extra carbon dioxide and the deficit in carbon dioxide eliminated in the periods of voluntary change and of recovery, respectively, were exceedingly small, as illustrated by the value of 4 cc per minute in the third group of periods on March 25.

TABLE 2. BALANCE OF CARBON DIOXIDE AND OXYGEN: ABNORMAL PERIODS VERSUS RECOVERY PERIODS

March 25, 1935				April 3, 1935			
Kind of period	Duration	Carbon dioxide	Oxygen	Kind of period	Duration	Carbon dioxide	Oxygen
	minutes	*cc*	*cc*		*minutes*	*cc*	*cc*
Deep breathing	5	+485	+159	Deep breathing	5	+754	+198
Recovery	30	−420	−156	Recovery	30	−763	−352
Shallow breathing	5	−75	−71	Deep breathing	5	+594	+193
Recovery	30	+30	+9	Recovery	30	−583	−242
Deep breathing	5	+265	+74				
Recovery	30	−375	−126				
Shallow breathing	5	−200	−171				
Recovery	15	+225	+162				

The balances between the extra oxygen absorption and the negative oxygen absorption in the recovery periods were not so good, although there were two (in the first and the last groups on March 25) that showed excellent agreement. Here again, however, the differences in the balances of oxygen in the two groups of observations on April 3, calculated on the basis of per minute figures, are only 5 cc and 2 cc per minute. It might be assumed that the increases in oxygen absorption during the deep breathing periods were due to the exercise of overventilation. However, the study of Liljestrand[7] and more recently that of Nielsen[8] on the effect of overventilation on the oxygen consumption per minute give values so low (0.3 to 0.7 cc) per liter of extra ventilation that this exercise can in no way account for the extra oxygen absorption. Furthermore, if the extra oxygen absorption were due to excessive activity in the effort at ventilation, this would not explain the lower values of oxygen absorption in periods of shallow breathing, nor the compensation values in the recovery periods following either deep breathing or shallow breathing. It is apparent that there is not only an extra elimination of carbon dioxide in overventilation, but also an extra absorption of oxygen.

This study shows that the human organism requires a significant amount of time for readjustment and compensation, when changes are produced in the equilibrium of its gases by voluntary over- or underventilation of the lungs. Alterations in breathing comparable to these voluntary acts are produced by emotional states and by slight amounts of activity. Once the organism is disturbed, it immediately attempts to regain equilibrium. Robertson[9] has stated that life is one constant struggle to regain equilibrium, and Cannon[10] has shown that a number of mechanisms aid in maintaining homeostasis. However, equilibrium is not regained instantly but requires time for its completion. The continuance of life depends upon a nutritive supply, but the ingestion of food immediately brings about a disturbance of equilibrium in the organism. In the study of the processes that go on in the body after the ingestion of food, the determination of the respiratory quotient is of value, as it is an indication of the kind of food burned and the rapidity with which the combustion changes from one type of substance to another. As the food supply varies with different animals, the broadest knowledge concerning the processes that take place in the body can be obtained only by studies on different species of animals. Such investigations have been made possible through cooperation with other departments of the Carnegie Institution of Washington and with laboratories outside the Institution. The effects of ingestion of simple sugars have been determined with the primates (*Rhesus* monkeys) of the Department of Embryology and with the normal and dwarf mice of the Department of Genetics, and similar determinations on goats are in progress in cooperation with the Laboratory of Animal Nutrition of the University of New Hampshire. A study of the factors that play a rôle in the utilization of sugar by man has had the cooperation of the Harvard Fatigue Laboratory. A study of the peculiar metabolism of protein by the Dalmatian dog has been made jointly with a member of the Department of Biochemistry of the Harvard Medical School. All these researches have dealt with the investigation of the respiratory quotient, the index of the kind of food burned in the body, and the rapidity of the changes in the bodily processes after food ingestion (storage of material and increased combustion) when the organism is striving to bring itself back to a condition of equilibrium. These cooperative studies on various animals have been made for the express purpose of aiding in the interpretation of results obtained on man.

---

[7] G. Liljestrand, *Skand. Arch. f. Physiol.*, vol. 35, p. 199, 1918.

[8] M. Nielsen, *Skand. Arch. f. Physiol.*, vol. 74, p. 299, 1936.

[9] T. B. Robertson, "Principles of Biochemistry," 2d ed. Philadelphia and New York, 1924.

[10] W. B. Cannon, *Physiol. Rev.*, vol. 9, p. 399, 1929.

FORM, FORCES, AND FUNCTION IN PLANTS

H. A. SPOEHR

Chairman, Division of Plant Biology
The Carnegie Institution of Washington

> It is the intuition of unity amid diversity which impels the mind to form science. F. S. Hoffman

In the old university town of Uppsala, some two hundred years ago, there was hatched a scheme which for daring and scope had rarely been equaled. It was one of those presumptuous, Utopian plans which evolve only in the unfettered mind of youth, a plan demanding more imagination and of vastly greater magnitude than any conceived by Gustavus Adolphus or Napoleon. In many an older head similar aspirations had been cherished, but most of these had been quenched by the very magnitude of the undertaking. It was given to two young, impoverished students to conceive and shape the principles of the scheme. As a first step toward its realization they divided between them the three great kingdoms of the earth—the mineral, the vegetable, and the animal. Their plan was to order, arrange, and name all there was, and all there was to be, in these great domains, according to their *systema.*

These zealous youths were Carl Linnæus and Pehr Artedi, by paternal design respectively students of medicine and of divinity. But fate overrode parental ambitions; Artedi was drowned in a Dutch canal, and it was left to Linnæus, though he finished his medical studies, to carry out and elaborate the great plan.

Observant men had long been impressed with the tremendous diversity of nature, particularly among the living things. The multiplicity of form taken on by living organisms seemed to present an almost hopeless wilderness to the orderly-minded. The great voyages of exploration had increased these domains apparently beyond any hope of survey; from the Americas, Asia, Africa, from the teeming tropics and the sparse deserts new and strange creations appeared. There never was a period when the need for order and the urge to arrange were so perfectly timed. Out of these efforts grew a universal language, the elements of which had been envisioned by Linnæus and Artedi. And the magnificent feature of the scheme of these two youths was not only that it was, in fact, set into motion and pursued by Linnæus during his entire life, but unlike all the conquests of the Napoleons, which so rapidly fell into ruin, this youthful endeavor prospered, and grew with altering needs long after its guiding spirit had gone to his rest.

Looking back over two centuries of restless activity of this army of collectors, arrangers, and classifiers, of this ceaseless effort to discover some thread which binds like to like amid the tangle of multiformity, one is impressed not only with the enormous industry and persistent application shown, but even more with the stupendous compass of the domains concerned. The urge to systematize and to order penetrated all substances and forces of the earth. From this effort sprang separate disciplines devoted to particular fields or kinds of substance. Because of limited diversity some fields were more rapidly arranged and their tillers could turn from classification to measurement, from mere ordering to the determination of the more exact arrangement of the parts. "Classification is a half-way house between the immediate concreteness of the individual thing and the complete abstraction of mathematical notions."[1] Thus much of early inorganic chemistry was devoted to efforts at classification, with the great reward of the periodic system and its amplifications. Out of the early efforts at forming types, organic chemistry rose to be an exact science. But these were relatively simple tasks as compared with the classification of living beings. The number of "species" or units of unorganized matter was relatively small, and, moreover, it could all be done within the comfortable confines of a laboratory. The living organisms offered far greater diversity and had to be brought in from the far corners of the earth, from the air, from the waters, and from the waters under the earth.

In this short period of time, short as intellectual efforts must be measured, there has been wrought at least a definite semblance of order. And most of this has been accomplished since Linnæus set out on his great venture. No one would be so rash as to claim finality for any of these arrangements. But they have given us a remarkable working basis for the further steps. Here we have the count of this task, naturally in approximate figures:

---

[1] A. N. Whitehead, "Science and the Modern World." Cambridge University Press, 1928.

92	elements
1,500	minerals
30,000	inorganic compounds
300,000	carbon compounds
1,500,000	species of animals and plants

It stands as an eloquent expression of a faith, an intuitive faith that there is order in nature, which is the impelling motive of science. This faith has been rewarded with at least a measure of understanding and, stimulated by vivid imagination, has sustained the ever quickening labors to penetrate the unknown.

The tremendous diversity of living beings and the fact that many of these were confined to particular and often inaccessible parts of the world, together with their complexity of form and function, imposed upon the study of plants and animals a much heavier burden than was the case in the efforts to classify the elements, the inorganic compounds, and even the carbon compounds. Yet this very diversity of organic nature seems to have offered opportunity for the discovery of fundamental phenomena not so easily detected in the simpler systems. Ideas of the fixity and permanence of the elements grew with the advancement of the science of chemistry, and the immutability of these units was long regarded as an established fact. The lack of permanency in nature and the concept of mutability were recognized in biology long before they were in chemistry. Ideas of organic evolution penetrated the thinking of biologists and were considered as an important phenomenon in organic nature long before such concepts were advanced in chemistry through the discovery of radioactivity and of isotypes. But the discovery of transmutation of the elements was followed rapidly by an understanding, or at least a very acceptable theory, of the mechanism which was involved. On the other hand, biology can claim little more than that it has described evolution, and in a measure reconstructed the steps through which it has passed. The means by which these changes occur in the form of living things remains one of the greatest problems of science and of all humanity.

Form

It is thus generally recognized that the diverse forms of animals and plants, as found in the world today, represent one stage in the history of organic evolution. The arrangement of these forms in such manner as to display their natural interrelationship and to aid in visualizing their stage of evolution has become an outgrowth of the early efforts at classification. Systematics, through the differentiation of organisms on the basis of their morphological characters, is one of the most time-honored methods of unraveling the apparently limitless variety of forms which are now found in nature, and of others of which we can discover only the lifeless remains. The theory of organic evolution has as its outstanding contribution the demonstration, so far as this is humanly possible, that these organisms have arisen gradually, and that many of the changes in their form have occurred as a result of influences which continue in operation and which are therefore discoverable. The discovery of the causes underlying these changes and the manner in which they come to expression, even perhaps the criteria to be employed for recognizing them, constitute a battlefield of biological research which promises to engage every arm of the service. No single approach can of itself triumph in what has already become a very involved and extended task. Rather is it evident that only by means of different approaches and the concatenate efforts of various disciplines can concepts be formulated which are adequate to explain the diversity of the phenomena encountered. These are essentially problems of primitive nature, phenomena which have been proceeding since the appearance of life on the planet, in a series of constant changes, irreversible but often recurrent.

Fortunate the man who can go directly to nature for an answer to his questions. Early in the organization of the Carnegie Institution of Washington it was recognized that the western portion of the American continent offered unusual opportunities for studying life in its native environment. Great areas of virgin country had been made accessible by modern transportation, but had not yet been altered by the devastating hand of man. A tremendous range in climatic conditions is here reflected in a flora and a geographical distribution thereof which are extraordinarily varied. The diversity of the scene, the multitudinous elements of which it is composed, and the complexity of interacting forces on living organisms all contributed to the formulation of different concepts and viewpoints as to underlying causes. It could not be predicted which viewpoint or approach to so complex a phenomenon would lead to a clearer understanding of nature. It is most fortunate that it has been possible to attack the fundamental scientific problems by means of different approaches simultaneously rather than to pursue one viewpoint and then another by a seriate attack as it were.

Under support of the Carnegie Institution of Washington, at several localities various problems relating to the influence of environment on form and distribution of plants have been intensively pursued, advantage being taken of the special environmental and vegetational characteristics of the particular region. Thus in the arid regions of the Southwest, in the Coast Range and Sierra Nevada of California, in the areas occupied by the great redwoods, and on the Great Plains, the influence of geographical and climatic factors was mapped, analyzed, and subjected to experimental study. Although the particular plant organisms and environmental complexes which herein were made the subject of special studies exhibited great variation, and the viewpoints and approaches of the particular investigators differed in many respects, it has become evident that the fundamental problems involved have many elements in common and are amenable to resolution on the basis of recognized scientific principles. The realization that these problems can be subjected to intensive, albeit extended, experimental investigation, with the aid of more recent developments in genetics, plant physiology, and biochemistry, has resulted in an appreciable unification of outlook in what were at the outset quite divergent approaches. If some of these viewpoints and approaches, often apparently conflicting and even contradictory, have served largely to call attention to omissions and oversights in others, they nonetheless will have been of real value.

Since the founding of the modern science of taxonomy by Linnæus, living beings have been classified on the basis of the form of certain of their parts. In this, the comparison of the visible and also some of the more minute parts has been used as a means of differentiating and classifying organisms into categories.

Of the approximately one and a half million species of living beings which have been thus recognized, about 133,000 are flowering plants. The general classification of a great many of this number has been fairly well worked out, and the "species" has been generally accepted as the central unit. But many of these so-called species, upon more careful examination, show differences, polymorphous characters, which in themselves are perhaps not sufficiently significant to warrant further differentiation by name into other species. The species may thus be composed of further recognizable units constituting subspecies, varieties, races, biotypes, or combinations of ecological forms. These intermediate or weakly differentiated forms are the "small species" of refined taxonomy. They constitute the material for much of taxonomic controversy. But of far greater importance is the fact that they may be the ray of hope for the student of evolution. The careful study of such polymorphous groups of very wide distribution may lay the basis for an understanding of the changes which have occurred in the processes of evolution—of differentiation from the ancestral stock during geographical dispersal or migration, and through the lapse of time. This, very briefly, has been the approach to the investigation of the problem of diversity in nature which has been developed in the Division of Plant Biology.

For the development of this approach to the general problem of organic evolution, various converging lines of evidence have had to be followed. It has first of all been necessary to make a careful selection of a number of polymorphous species of relatively wide geographical distribution, and this could only be done on the basis of thorough experience with the groups concerned. A knowledge of species composition, of the slight discrepancies and deviations in form, is obtained by comparing the taxonomic characters of plants collected from different parts of the world.

In those great dusty storehouses, the large herbaria, where lie deposited the amassed, lifeless evidence, from all corners of the globe, of the plant collectors' passion, this evidence of species composition is conveniently assembled in neat, yellowing folios. Long the victims of the illusion that discovery and classification is an end in itself, these collections are beginning to serve a more living god. They form the beginning of the scrutinizing search for variation in form, for some sign of evolution in progress. Also they give some picture of the extent and variability of the units.

But they are after all only samples taken at random and often without the discerning eye, and therefore do not give a complete picture of the variations existing in nature. To discover that, it is necessary to go directly to the field, and on the basis of extensive collections over wide areas of differing climate to compare, correlate, and finally to examine by means of other than purely form characters. These herbarium and field studies have revealed a remarkable diversity among some of the taxonomist's species. Again other species have been found to exhibit a striking uniformity of composition throughout a wide range of geographical distribution. But for the present purpose, most important are those which are composed of almost innumerable variations, whose characters seem to pass into neighboring species through gradations of form. For example, a California weed of wide distribution, *Madia elegans*, has been found to contain several hundred such units of slightly different form, and other examples have been found in which the interplay of obscure genetic and ecologic factors results in even greater morphological complexity.

It is, therefore, not surprising that efforts to establish principles of classification of these complex groups of organisms should result in serious controversy, and that some doubt should have arisen as to the very existence of natural systematic units. In order to extricate himself from these difficulties the taxonomist has finally taken recourse to a rather fundamental change in approach to his problem. Instead of being content with purely observational data he has adopted the experimental method, and has taken the first steps toward the founding of a new discipline of experimental taxonomy.

Science itself is a marvelous example of evolution. No new development of scientific thought is a product of spontaneous creation; it has a long line of parents and grandparents, a period of gestation when it is unknown except to a few hopeful souls, and its birth is heralded by doubts and headshakings by all but the few who have mothered it. A new scientific discipline is usually the offspring of two sympathetic modes of thought which are in need of mutual help. It bears the characters of both parents, and until it proves its own power is pointed at as a bastard by the older generation. Physical chemistry, biochemistry, cytogenetics, and so on, in their first generation struggled with their own inferiority com plexes, and attained to a respectable position in the society of science only through the demonstration of their ability to serve useful ends. In like manner experimental taxonomy came out of classical taxonomy, sired by genetics.

One of the main objectives of experimental taxonomy has been the establishment of a more objective delimitation and classification of species than is possible on the basis of the conventional characters used by the taxonomist. Without an understanding of the significance of each of the parts of a species complex it is manifestly difficult to determine which of these may play a rôle in evolutionary change. Such experimentation is of necessity a slow process and involves masses of data. Yet as the results from each annual harvest have been assembled and examined, a definitely clearer concept has been gained of the evolutionary significance of the various orders of systematic units such as species, subspecies, and varieties. Too wide generalizations can, of course, not be drawn; they apply to the particular group of plants which have been subjected to this form of experimentation. The concepts are of special importance, as their descriptions serve as the definitions of the systematic units which the experiments have disclosed. Hybridization has been found to be a particularly useful experimental test to differentiate the array of forms constituting a polymorphous "species" such as is here being considered.

It has been concluded by the investigators conducting these researches that these forms (or groups of similar individuals) are separated by mechanisms which are inherent in their own structure. Thus true species are separated by genetic barriers or incompatibilities. Their hybrids are either partially or completely sterile according to the degree of incompatibility. In extreme cases species are so incompatible that no fertilization takes place; in others fertilization is possible, but the development of the first generation hybrid stops in the early embryonic stage. Others have seemingly good seed which, however, does not germinate; while still others yield hybrid seedlings which are dwarfish, slow-growing, or which die before maturing. Finally, some species can form vigorous and normally flowering hybrids in the first generation, but they are entirely or almost entirely sterile; or sterility appears in the next generation. The incompatibility is expressed in many ways and gradations, but in the first or succeeding generations abnormalities and loss of vigor in one manner or another become apparent, so that finally a very large percentage of the hybrid product will be wiped out because, from lack of vigor, it is unable to compete in nature. The ramifications of this manifestation of incompatibility in interspecific hybrids is very great and the detailed cytogenetic elucidation thereof constitutes in itself an intricate study. For the present purpose it must suffice to emphasize that species represent a stage of evolutionary process "at which the once actually or potentially interbreeding array of forms becomes segregated in two or more separate arrays which are physiologically incapable of interbreeding."[2]

By similar means it has been found that what are termed *subspecies* or *ecotypes* produce hybrids which are fertile and vigorous in both first and second generations. There are no apparent genetic incompatibilities. They appear to show geographic or ecologic preference, and the characters by means of which they are distinguished probably have selective value in their natural environments. Lacking genetic incompatibility, they keep pure only through their geographic or ecologic isolation. Where they meet in nature and cross, purity is probably attained by natural selection.

Varieties also are completely interfertile and give rise to fertile and vigorous hybrids. But they are neither geographically nor ecologically isolated, and the characters by means of which they are distinguished appear to have no selective value in the environment in which they live. Their occurrence is usually spasmodic and several varieties may be found in the same geographical locality. They are, as a consequence, of little evolutionary or taxonomic interest. Similarly, the recessive "mutants" so commonly used in genetic investigations appear to have little evolutionary importance. Because of their inferior viability and vigor they are seldom found in natural populations of self-sterile species, but they are segregated when individuals from natural populations are inbred.

FORCES

One of the most significant features of plant life is that each individual is permanently bound to the environment in which it starts life. This is more particularly true of the higher, flowering plants. From the time of the germination of its seed till its death the plant's entire existence is spent in one spot; it is there exposed to all the elements and changes of season. It is self-sufficient in that it is capable of manufacturing its own food. If it can withstand the forces of weather, if it can secure the necessary raw materials for its manifold functional activities, it will be able to survive. Between mere survival and luxuriant development are innumerable stages in all of which the forces of the environment impress themselves upon the development of the plant.

---

[2] T. Dobzhansky, "Genetics and the Origin of Species." New York: Columbia University Press, 1937.

These forces of the environmental complex are constantly changing in relative intensity and effectiveness. No two hours, days, or seasons are an exact repetition of previous ones, no two spots on the earth are exactly the same so far as the plant is concerned, and with greater changes in location corresponding modifications in environment exist. Most plant migrations are effected by means of seed dissemination, by which means the offspring is always carried into an environment differing more or less from that of its parents. To some of the changes in the intensity of environmental forces the plant is indifferent; others may affect its life profoundly. Also some plants are extraordinarily sensitive to very slight modifications in their environment, while others are able to exist over a wide range. Examination of the diversity of form taken on by plant organisms clearly suggested that form was an expression of the forces in the environment to which the organism was exposed. More critical and minute study of variations in form as used in classification, where form is considered in terms of development, has given further indication of the influence of changing forces in course of time.

What applies to the form of plants in relation to the environment applies equally to their function; in fact the three, form, forces, and function, are so intrinsically integrated in the living world that no one of them can be adequately considered without regard for the others. Life is essentially a dynamic concept; the organism is characterized by functional organization, operating in an environment which, because of its inherent physical and chemical nature and because of the influences of organisms within it, is itself dynamic. The organism cannot be regarded simply in the light of a highly organized machine operating within a physical environment, or as a thing set apart from the inanimate world. It is too inti mately a part of this environment, too intimately interlinked with the physical forces of the environment, to be so regarded if a comprehensive concept of nature is to be obtained.

Of course, so complex a problem can for practical reasons be handled only by attempting to resolve the general problem into a number of simpler ones by means of analysis, with the expectation that each one of these can be more easily dealt with individually. The attempt is made to analyze the complex situation into the various factors and elements which are operative, to study these in isolation, in order to penetrate into the intimate workings. This analytical mode of approach to a complex problem, this resolution of the intricate into its constituent parts, and investigation of these in isolation, constitutes a fundamental principle of the modern scientific method and is essential in order to gain an understanding of nature. But the important feature of the method is the recombining of these parts so as to integrate them into a functioning whole. In fact this is the ultimate aim of the method and constitutes an essential correlate of analysis.

Ecology. Incomplete as such syntheses must of needs be at this stage, the ecological view of nature has served to present a much more comprehensive picture of the living world, of the biotic community in which the inanimate materials and forces, as well as the plants and animals, are interlinked in a dynamic whole. Moreover, just as plants and animals can be divided into discrete groups or species representing a stage or time level in the course of evolution, so whole biotic communities can be recognized as representing a stage in an ever changing complex of organism and environment. Grassland and forest, sand dune and lichen on bare rock, represent stages in a succession of changes. For the moment each such community seems discrete and of enduring stability, but when viewed more intimately and with sufficient perspective, it is seen to represent delicately adjusted dynamic equilibria. Each form carries the germ of its own destruction, long imperceptible and kept within check, then gradually giving way to a series of changes and yielding to some other complex of forces, finally to take on an entirely different aspect.

The Great Plains and adjacent regions of the western United States have during the past generation offered unusually valuable and dramatic evidence of the interaction of organism, including man, and the natural environment. The rational use of these great land areas and the avoidance of recurrent tragedy in terms of human happiness depends in no small measure upon the exercise of scientific foresight rather than upon political expediency. The basis of this foresight has in a large measure been laid through the extensive ecological investigations of this region. These have clearly established that in nature change is universal and inevitable; that nature's way is to meet change with change. The organisms of the past, both individuals and communities, that have remained static have passed to extinction, and only their lifeless remains are left to record their non-adaptability. The recurrence of conditions both favorable and unfavorable to certain groups appears to be a characteristic of these changes. Man has reached the stage of being able to record and collate the effects and evidences of the dynamic universe of which he is a part. The mastery of his destiny depends in no small measure upon his capacity to compel his own actions to harmonize with this accumulated knowledge.

The desert. Among the first projects undertaken by the Carnegie Institution of Washington was one devoted to the investigation of life under extremely severe environmental conditions. The story of the emergence of life from the sea had been written, at least in outline; and while innumerable problems still remained and were, in fact, multiplying in hydra-headed fashion, it was realized that many novel lines of development were presented by land plants. In these there had been perhaps no more differentiation with reference to reproductive processes than occurred in marine organisms, yet the land plants showed vastly greater diversification in the vegetative or non-reproductive features. It was realized, moreover, that scarcely less striking than the attainment of the land by plant and animal life has been its gradual colonization of the drier and drier parts of the earth's surface. The movement of the biota from moist lands to dry ones or their adaptation to changing conditions in the same place has resulted in the development of striking structures in plants and profound changes in their life habits. These remarkable effects in plant life are most strikingly exhibited in the extraordinarily severe climatic conditions of the desert areas. Such areas, or semiarid ones, constitute one-third of the earth's land surface and nearly one-fourth of the area of continental United States. In terms of physical environment such as extremes of temperature, of lack of water over long periods of time, of intense sunlight, parching winds, and cold clear nights the deserts present conditions which represent the extremes for the survival of life. Yet this delicate, sensitive stuff called protoplasm, which is the carrier of all life activities, has proved to be more stable and persistent than the dark mountains and the very rocks among which the plants of the desert grow.

The enigma of the desert is its life. Our notions of living things, their form, how they function and their relations to each other, etc., have for the most part been formulated from experience and observation in temperate or tropical regions. The contrasts which one first encounters on the desert are startling and harsh. But the true depth of these contrasts and their objective significance reveal themselves only gradually through intimate living in the environment. Whence came these strange plants and animals which inhabit the desert and how are they able to maintain their life, to adjust their functioning to the extreme physical conditions to which they are exposed? These are the recurring questions; for these organisms have retained the functions essential to all higher plants, such as the manufacture of their own food, growth, and reproduction, by apparently an astounding capacity to modify some of the ways of their ancestors and to become organized for the environment in which they find themselves.

It was essential, first of all, to obtain a picture of the vegetation of the arid regions on the background of the complex of environmental conditions of which it is a product. This has been done by means of extensive exploration and study of the composition, relationships, and distribution of the desert vegetation. It is now apparent that in many respects the plants of the desert are less profoundly influenced by one another than are those of moist regions, and that they are more closely affected by the environment. But the interrelations between the life and the environment are intimate and intricate. Every aspect of the desert is closely tied to every other aspect of it. The influence of the climate on physiographic and soil processes, the relation of these to the vegetation, the influence of vegetation on the life of animals, the manifold influences of the animals on plants, and the reciprocal effect of plant and animal life on physiography and soil, all combine to form a vast nexus of relationship which demands a sharp and comprehensive view if the picture is to be true. For a visualization of the features which are truly characteristic and significant it has seemed important to make comparative studies of other arid regions, and this has been done throughout North America, and in Algeria, Libya, Australia, and South Africa. Identical or closely similar features in structure, function, and adjustment to environment are found again and again in far-separated regions or in different continents. These universal features are definitely related to the ways in which the evolutionary processes have molded the vegetative organs of plants into successful desert forms. The features which are distinctive to particular desert regions are the characteristic plant communities, the regions from which they have been derived, and the ancestral stock from which they have descended. It seems clear that, while the origin of the desert floras is primarily a floristic problem, this is intimately bound to causational ecology. In an attempt to trace their origins, the life histories of the plants, their relation to the environmental complex, and their distributional characteristics are quite as important as are their taxonomic relationships.

The desert's physical environment itself presents an extremely complex picture, especially when this is regarded in its possible relation to the life of the region. Not only do the various climatic factors, such as rainfall, evaporation, and temperature, vary over a wide range, from those that are extremely unfavorable for organisms to those that are merely difficult; but the distribution of such climatic factors during the course of a year, for example, is quite as important for survival as their intensity. The long and often irregular periods during which the plants are exposed to extremes impose rigors which are definitely determinative for survival. A true conception of these physical conditions can be gained only on the basis of extended and painstaking measurements over wide areas. The accumulation of such records and their correlation with the life habits of the organisms alone can form the true picture of the environmental influences. For example, the ratio of rainfall to evaporation and the manner in which water is held and supplied to plants by soils have been found to be highly important in controlling plant distribution and in determining the limits of the desert, the line between grassland and forest, and the range of certain forest types.

The extensive experience, now extending over thirty years, gained from the study of the relation of climate to organism has clearly demonstrated the fundamental importance of time in these relationships. A single season or year, or even a few years, of observation can yield only a fragmentary picture, and often a quite erroneous one, of the interplay of climatic factors on living organisms. Whether the variations in climate are truly cyclical or whether these only approach cycles in the astronomical sense, as well as the definite correlations between solar and terrestrial phenomena, obviously will require further study and elaboration. The fact is nevertheless clearly established that there are recurrences of climatic conditions favorable and unfavorable to the development of plants. Moreover, it is now established that the great mass of observations on the relationships of plant and animal life to these environmental conditions is definitely amenable to scientific collation. The concepts arising from such investigations will ultimately have real social significance and be of human service. They constitute the only rational basis for a solution of the problems which are presented in the management of the half-billion acres of arid and semiarid range land in the United States, and the larger areas in Africa, Australia, and South America.

Transplant experiments. The significance of characters of form for the classification of organisms has already been alluded to. The manner in which, through the ages, the forces of the environment exert their influence upon organisms, resulting in groups which are apparently adjusted to their surroundings, is one of the most fundamental problems of biology. There appears to be little doubt that most plants exhibit modifications of a more or less obvious nature according to variations in the climatic complex to which they are exposed. It is, however, among the first necessities for rational classification to be able to distinguish between fixed, or hereditary, characters and possible temporary modifications resulting directly from environmental influences. The distinction between fixed and modifiable characters can be made only on the basis of actual experiment. This has been one of the main objectives of the investigations in experimental taxonomy which have already been mentioned, and the experiments directed toward these objectives have come to be known as transplant experiments.

The principle of this type of experimentation is very simple.

When it is desired to test the permanence of characters of related forms which occur in diverse habitats, the plants are moved into the same environment. For example, if the forms appear to vary with altitude, as many of them as possible are brought together at a station of medium elevation. If the differences disappear, it is safe to assume that the original forms were nothing more than ecological modifications. If, on the other hand, each form retains its characters for a number of years, it is assumed that the differentiating characters are hereditary. Conversely, in order to induce variation of form and to test the degree to which characters may be modified by environment, very different operations are required. For such purposes a type of experimentation which is virtually the reverse of the one just described is followed. That is, instead of moving diverse types into a single environment, a single plant, or, more accurately, parts of the same plant, are moved into diverse environments. Induced changes of form of whatever nature are of interest, but the center of interest lies in those changes which parallel the variations found in nature. Various modifications and ramifications of this type of experimentation have, of course, been developed, but these are the essential features. The different environments are obtained at experimental stations of the Carnegie Institution located along a transect extending from sea level on the Pacific to timber line in the Sierra Nevada of California. To these the divided individuals are transplanted.

Various groups of plants have been worked with, and for the most part these have been perennials in order to avoid complications due to heterism. It is already evident that no two groups of plants behave in exactly the same manner. What appears to be the same character may be highly modifiable in one species, but unchangeable in another.

The conclusions of special importance drawn by the investigators in this field are that every plant retained its individuality irrespective of the environmental conditions to which it was transplanted. A few plants exhibited quite spectacular changes in vegetative characters when grown in different environment (altitude), but these modifications never obscured the individuality of the plant.

Every transplanted plant has a certain capacity for modification when grown in a different environment. These modifications vary with different species, with races, and even with individuals of the same race. The order of magnitude of these modifications is so small, however, that ordinarily no confusion is caused in the taxonomic identification of well-established species or varieties. These modifications are apparently temporary, quickly induced, and reversible.

Plants as recorders of climate. While, then, we are as yet unable to give a completely satisfactory interpretation of the manner in which the environment has molded plants in their evolutionary development, there is no doubt that the environmental and climatic influences on form and function are clearly discernible. As a consequence, plants can also serve as indicators or recorders of past climatic conditions and they thus constitute a most valuable means of extending our information about such conditions on the earth backward through historic, prehistoric, and even geological time. The earliest records of the effects of climatic changes are the fossil plants, which in conjunction with geological evidence constitute one of the most valuable means of determining the nature of climatic conditions existing on the earth millions of years ago. A more complete account of some phases of this subject is contained in another section of this book.

For a record of the climatic changes in more recent times, but running back as much as several thousand years, we must look to the massive plant structures of great age. Of these, trees are the most useful, because of their capacity to form an annual accretion of wood each year. The size of this annual ring of wood and its structure have been found to be clearly influenced by the climatic conditions under which it is formed. Thus in certain localities, and in certain species of trees where water supply is the determining factor in the growth of the annual ring, it has been possible to establish long chronologies depicting the available water and the cyclical nature of the fluctuations in this factor throughout historic times.

From the very extended investigations of the manner in which the forces of the environment impress themselves upon the structure and functioning of the plant organism, it has become evident that the record thus produced is an exceedingly complex one. While the extremes of one factor or another may give clearly discernible effects, under other conditions the interlinking of several factors may yield less easily perceptible correlations. As a consequence, a vast amount of intensive research is still needed to form the basis for interpreting the many structural effects produced in plant organisms by the various factors of the environment. As an example of the sensitiveness of the plant cell to relatively slight changes of environmental influences, and of the way in which these are permanently impressed upon the cell, may be cited the recent observations of Dr. I. W. Bailey and his collaborators. They have found that such delicate and minute structures as the cellulose walls which surround each cell in the living portion of the tree are composed of material of alternately greater and less density, and that these variations apparently correspond to daily fluctuations in some environmental factor or complex of factors such as water supply, temperature, and light. In much the same manner, as minute a structure as a single cotton fiber reveals the fluctuations of the environmental complex to which it was exposed during the time it was being formed as a product of cell metabolism. These structures simulate on a microscopic scale the grosser tree rings, and record the relatively slight daily fluctuations of the environment as compared with the more drastic seasonal changes recorded by the tree rings.

Function

On the foregoing pages attention has been directed to the efforts of plant science to arrange and order its material by means of classification. Some indication has been given of the basis upon which this systematization has progressed, of the influences which contribute to the great diversity of living organisms, and of the way in which the purely observational efforts of classification are giving way to experimental ones. In all these systems of classification the major emphasis is upon differences in form and in the structural elements of the organism. Fundamental as these efforts are, they represent but one approach to a larger task. These multitudinous forms carry out certain functions which are common to them all. The manner in which this is accomplished has in a large measure been determined by the environment in which the plant is placed, but basically the requirements for life activity and the manner in which this manifests itself are much alike. As a consequence, the functioning of plants has been studied from quite a different viewpoint from that followed in the efforts at systematization. Instead of stressing the differences and peculiarities which distinguish different forms of plants, function has been regarded largely as a property which is characteristic of all plants. Especially is this true of the higher, green plants.

The most significant of such common functions is perhaps that all living beings, just as all man-made machines, require a source of energy to maintain their activities. For the running of an engine combustible fuel is required. The maintenance of the life activities of a plant or animal organism requires a supply of fuel in the form of food. The energy from the fuel for the engine is derived through the burning of the fuel, that is, by its chemical combination with the oxygen of the air. The energy from food is obtained by the organism through an analogous though more intricate process. But in general principle the analogy between organism and engine holds. In all this the green plant exhibits one very important difference: it can manufacture its own food. All animals are entirely dependent upon the environment for their supply of food; so, of course, also is man, and so also does he take from the environment the fuel which runs his machines. To the extent that the green plant alone is capable of manufacturing its own food, it is a far more independent and self-contained being than any animal. Yet the plant, in turn, is dependent upon its environment for the raw materials out of which it manufactures its food and for the energy which it uses in this process.

By means of this process of food manufacture through photosynthesis the plant lays by regularly a small reserve for use during periods of stress and adverse conditions. It apparently has definite regulatory devices which prevent the complete consumption of its daily earnings during fair weather. Survival of the individual plant depends upon an available supply of reserve food for adverse times; survival of the species depends upon passing on to the next generation a patrimony in the form of food material, to enable the young plant to pass successfully through the critical stages of its early life to the time when it is itself an independent photosynthetic organism. All animals, including man, owe their existence to robbing the plant of these reserves. These reserves of the plant, either directly, or indirectly through first having nourished an animal, constitute our sole means of subsistence. Similarly our industrial needs for fuel, both coal and oil, are met by drawing upon the remains of plants from past geological ages.

These rapidly vanishing supplies we regard as our most important natural resources. In fact they are but a speck, almost an accident, in the great interlinking of life and environment on this planet. The whole environment, animate and inanimate, is in fact inextricably involved in man's well-being. Not only the consumable resources are involved in this, but quite as much the end products or results of man's activity. Life cannot be regarded only from the viewpoint of what is required from the environment for life's maintenance; of equal significance, especially for future generations, is what this life is doing to the environment.

To digress for a moment with an illustration which may not be without some social implication: To the biologist the phenomenon of hardship in the midst of plenty is not uncommon, since he has learned to consider not only a single need of an organism, such as food, or the organism alone, but rather the entire system in which the organism is active and its reciprocal relation to all factors which may affect its life. For example, the minute yeast plant has the capacity of using sugar as food. It works rapidly and, with adequate food supply, prospers and multiplies. The products of its activity are alcohol and carbon dioxide. These accumulate in its immediate environment and they are both harmful to life. If sufficient food is available, the organism continues its fermenting action, and the concentration of its metabolic products soon reaches a point which makes it difficult for the weaker members of the colony to live. With the continued activity of the more vigorous members, the metabolic products increase to a point which causes the death of a large proportion of the colony, until finally the entire community falls victim to the deadly influences of its own metabolic products. Thus, in spite of an ample supply of food and of the vigorous work of its members, the life of the community is lost, a victim of its own activity. The same phenomenon can be observed in organisms which form butyl alcohol from sugar; and many others, such as the lactic acid organisms, survive only when the products of their activity are removed by suitable means from the environment. Obviously the living organism may influence its environment, and consequently its own life, not only by taking materials from this environment, but quite as drastically by adding other materials to the environment which produce its own destruction, in spite of a surfeit of food.

In the first part of this article attention was directed to the tre mendous diversity of organic nature. It is not without interest that the approximately one and a half million different forms of living beings, all of which show some differences in life habit, are dependent for their existence upon a single chemical process, namely that of photosynthesis in plants. As an over-all process it is a relatively simple chemical reaction: carbon dioxide and water, under the influence of light, yield a carbohydrate, with the liberation of oxygen. This is the mother reaction for almost all life. Through all the great diversity of plant life, representing widely divergent evolutionary development, there is no evidence that this means of obtaining food has in any way been modified. With the exception of some of the colored bacteria, and possibly some of the algæ, all plants apparently follow the same course of obtaining their primary food. Within the body of the plants this primary food is changed into a great multiplicity of chemical substances. Herein the diversity of plants again comes to expression, for the variety of substances found in plants is enormous and differs greatly with the life habit and environment; differs, in fact, from species to species. Within the bodies of animals these carbon compounds are again altered, some broken down to simpler ones, others built up into more complex forms. The three hundred thousand carbon compounds, mentioned in our first table, include the host of substances found in plants and animals. Besides these, a very large number of carbon compounds has been synthesized in the laboratory. The great majority of these are directly or indirectly connected with the substances formed in the photosynthetic reaction of plants. That the plant produces many substances, other than food products, which may be of value to the manifold needs of industry is, surprisingly enough, only recently coming to realization, and quite naturally promises to effect some basic changes in the relation of agriculture to industry.

In order that the plant may carry on its photosynthetic manufacture it requires an adequate supply of raw materials and a source of energy to propel the process. These raw materials are carbon dioxide and water, which are derived from air and soil; the energy is the light from the sun. One of the remarkable features of the process is that the plant can utilize raw material which may be considered of such very low grade. In the case of land plants, which at least economically are of the greatest significance, they are dependent upon the supply of carbon dioxide in the atmosphere. However, only 0.03 per cent of the air is carbon dioxide. On this apparently thin thread hangs all the life of our planet. Under conditions most usually found in nature, if the carbon dioxide content of the atmosphere were to be reduced, the rate of photosynthesis would be correspondingly diminished; and, on the other hand, if, for example, the carbon dioxide content were doubled, the rate of photosynthesis would be increased approximately twofold.

Yet this supply of raw material upon which we must depend for our life's existence is not as frail a thread as would at first appear. The gigantic proportions of the earth's atmosphere constitute a reservoir of carbon dioxide, the magnitude of which is difficult to grasp. The resources of the air alone amount to about 2,200,000,000,000 (2.2×10^{12}) tons; and the amount contained in the sea has been estimated as fifteen to twenty-five times this amount. Some conception of the magnitude of these figures can be gained from a consideration of the fact that, in spite of prodigious quantities of carbon dioxide which have been poured into the atmosphere during the past fifty years through the industrial use of coal and petroleum, no evidence of any increase in the percentage of carbon dioxide in the air has been observed.

The driving force of photosynthesis is the energy, in the form of visible light, from the sun. The computations in the accompanying table may aid in an orientation of the magnitudes of energy which are involved in this process.[3]

DISTRIBUTION OF TERRESTRIAL ENERGY

Sources	Million calories per year	Calories per year
Solar radiation received at the outside of the earth's atmosphere	1,340,000,000,000,000	1.34×10^{21}
Energy used in the evaporation of water	340,000,000,000,000	0.34×10^{21}
Energy used in photosynthesis, on land, the benthos of the sea, but exclusive of plankton	162,000,000,000	0.162×10^{18}
Energy of all flowing water on the earth	50,000,000,000	0.050×10^{18}
Energy of the world's coal production	6,600,000,000	6.6×10^{15}
Utilizable water power	2,800,000,000	2.8×10^{15}
Utilized water power	80,000,000	0.08×10^{15}
Total work capacity of the human race	70,000,000	0.07×10^{15}

[3] H. Schroeder, Quantitatives über die Verwendung der solaren Energie auf Erden, *Die Naturwissenschaften*, p. 976, 1909.

Figures of this nature have value only in giving some indication of the total activity which is involved, of its relative significance, and possibly of the magnitude of our terrestrial system. Just as the size of a national debt gives little information as to the contribution which every citizen must make toward its liquidation, these figures give little indication of what is going on in the individual manufactories of the plant. And it is important to bear in mind that while the total output of this process is enormous, it is all done by piecework as it were. Each plant represents an individual, itself dependent upon a complex of environmental forces, each leaf of the plant a semi-independent factory of the system, and within each factory millions of units in which the process is actually going on. These units are microscopic organs, the chloroplasts, usually many in each single cell, so that an ordinary sunflower leaf, for example, contains about ten billion of these individual units in which the photosynthetic process occurs. So far as the photosynthetic process is concerned, the rest of the plant, the roots, the stem, and the pores, are merely ancillary organs, which supply the raw materials and carry off the manufactured product. The center of interest for the photosynthetic process is the chloroplast. The chloroplasts are as important to an understanding of this process of photosynthesis as the chromosomes are for heredity.

The chloroplasts constitute the apparatus in which the chemical reactions comprising the process are carried out. A knowledge of their behavior and composition is therefore of importance for an understanding of the mechanism of photosynthesis, by which we mean the exact description of the physical and chemical changes which constitute the process of photosynthesis. If we are to determine how a mechanism operates we must know its parts, of what it is composed, and its relation to the larger mechanism of which it is itself a part. Again, we can also reason from the manner in which the process is affected or controlled by its raw materials, and by the energy necessary to propel it, that such or such materials may be involved and that certain chemical and physical changes must play a rôle. During the past half-century both of these avenues of approach have been followed. Advances have been made in one direction and then in another as experimental techniques have been developed, and frequently the two means of attack on the central problem have supplemented each other. But the process is essentially a chemical change in which light energy is converted into chemical energy. For this reason particular importance attaches to its interpretation in terms of chemistry.

A characteristic of outstanding importance is the fact that the chloroplast is a living organ of apparently extraordinary sensitivity. The slightest injury results in cessation of its functioning. Its structure or perhaps the colloidal relationship of its chemical components is absolutely essential to its operation. This not only is true for the photosynthetic reaction itself, but has been found to apply to other reactions about which we are relatively well informed, and which also occur in the chloroplast. Certain enzymatic reactions which are known to take place in the chloroplast are completely checked when its structure has been disturbed by means which do not inactivate the enzymes. It appears, therefore, that the arrangement of the parts in a functioning structure is essential, and that this structure is very easily disturbed.

It has been possible to determine the nature of the more obvious parts of the photosynthetic apparatus by means of chemical methods, though an exact knowledge of these parts is still lacking, and some of the constituents have so far eluded detection. The chemical properties of the chloroplast constituents, and consequently the rôle they may play in the photosynthetic process, can in a measure be determined from their molecular structure. For this reason a very great effort has been made to isolate these substances in pure form and to establish their structure and their physical and chemical properties. These efforts have, however, been beset with many difficulties, because for the most part the components of the chloroplasts not only are extremely complex substances but are remarkably labile, so that they are apt to undergo changes in the process of isolating them from the plant.

Of the more obvious components of the chloroplasts we have at least a fair knowledge. The most striking of these is the green coloring matter, chlorophyll. Actually composed of two pigments, the fundamental chemical structure of chlorophyll is now fairly well understood. The more intimate study of the optical and chemical properties of this remarkable substance is revealing a most extraordinary adaptation to the rôle it plays in the diverse environmental conditions under which the plant must operate in nature. The finer aspects of the properties of chlorophyll are still to be worked out, and will require chemical and physical investigation of the highest refinement. This is particularly true because there is now little doubt that the chlorophyll which has been isolated from the plant and studied in the laboratory differs in some very significant respects from that existing in the plant. In this connection it must be stressed that all efforts to reproduce the photosynthetic process outside the living cell have met with failure.

Besides the chlorophyll there are within the chloroplast other, less obvious pigments, of very different chemical structure. These are the yellow or orange carotenoid pigments. Here is an example of having isolated a part of the machine without knowing just what rôle it plays in the functioning thereof. Although these pigments are present in every chloroplast, the hypothetical schemes which have been formulated for a mechanism of the reaction cannot with certainty ascribe specific rôles to these members of the apparatus. Investigation of the carotenoid pigments has revealed that they constitute a group of chemical compounds of much greater complexity and diversity than was at first supposed. Composed essentially of a long chain of carbon atoms, partially saturated with hydrogen atoms and capable of adding and releasing hydrogen, it is conceivable that they may play a rôle in the hydrogenation mechanism which is essential to the reduction of carbon dioxide to carbohydrates.

The fact that we can at the present not ascribe a definite chemical or physical rôle to a particular component of the photosynthetic apparatus is in itself no reason for discouragement. The struggle with this problem of the mechanism of the photosynthetic process has been under way sufficiently long to warrant the conclusion that victory is not going to come by means of a few spectacular and brilliantly executed sorties. It is manifestly a problem demanding the patient strategy of a hard siege. Each position must be secured, and in such a manner that it can be positively relied upon by all arms of the service. In terms of scientific research this means that the results must be attained with the greatest possible precision available under the circumstances. In the end, that means economy of time and effort. Time and again a promising avenue of attack has yielded only disappointment because of insecure or fragmentary data. Well-established and reliable information concerning the components and structure of the chloroplast, coupled with precise measurements of the materials used and formed by the reaction, including the energy changes, constitute the weapons which must be relied upon for a coordinated attack on this problem. Some few positions have been thus already fairly well established. These permit us at least to adumbrate some relationships of the various components. Such outlines constitute the hypotheses which have been formulated to account for the observed facts. As the latter come into clearer focus the relationships are also more clearly discernible, and this makes possible the devising of experimental arrangements for testing the hypotheses, and finally the formulation of more definite laws.

That the hypotheses regarding the mechanism of the photosynthetic process are still in a rather unsatisfactory state may be accounted for by the fact that there are still many essentials which have not been established with any degree of certainty, and these must therefore be treated on the basis of mere assumption. Every clearly established fact as to the chemical and physical composition and as to the behavior of the photosynthetic organism serves to remove our reasoning farther from the domain of mere speculation and to support it with sound chemical analogy. The following may serve to illustrate this. Manifestly the first step in the photosynthetic process is the absorption of carbon dioxide from the medium surrounding the plant. Yet until very recently no critical investigations have been carried out to determine the chemical nature of the mechanism or the substance by means of which this is accomplished. While there are still some features of this problem which are in need of elucidation, it has now been established that even unilluminated leaves have a definite carbon dioxide binding capacity. The mechanism responsible for this is of a reversible nature, so that the carbon dioxide thus absorbed can again be easily passed into the photosynthetic mechanism. In a large part the absorptive agents appear to be phosphates and the carbonates of calcium and magnesium. The more exact identification of the various components of the photosynthetic apparatus is thus making possible a correspondingly clearer picture of the elements which are responsible for this first step. To this extent they serve to elucidate the entire process.

Much valuable information for the formulation of ideas concerning the mechanism of the photosynthetic process has been obtained from quite a different approach. The energy requirements of the process are of great significance. This is true because the photosynthetic process is primarily an energy-storing reaction. It can be calculated, with a high degree of certainty, that the conversion of one gram-molecule of carbon dioxide into the equivalent of one molecule of carbohydrate (which essentially represents the photosynthetic reaction) requires energy equivalent to about 112,000 calories. It is assumed that in photosynthesis this energy is obtained from the light absorbed by the plant. Now, photosynthesis proceeds perfectly normally in light of all visible wave lengths, from the violet to the red end of the spectrum. However, one unit or quantum of radiation of red light does not possess enough energy to carry forward the reaction, and it would require about four quanta of red light to make this possible. Theoretically this means that the reaction would probably proceed in four steps from carbon dioxide to carbohydrate. Although reasoning of this nature can give us very little information concerning the exact chemical reactions which are involved, it does make possible a fairly clear delineation of the criteria which must be met in setting up any possible chemical scheme of the photosynthetic mechanism.

The chief value of the approach to the problem through energetic considerations of this nature is that, in a measure at least, they are amenable to experimental test. Experimental results obtained by this means may yield a framework from which the more precise chemical steps involved in the reaction can be envisioned. At the present time this approach is being developed with much interest, especially as there is indication that under certain circumstances the four-quanta yield referred to above can be attained. However, in these experimental tests we are not dealing with the relatively simple chemical systems from which the photochemical laws on which we are relying have been derived. The fact that the photochemical system in this case is a living organism, that, moreover, we are dealing not simply with a series of photochemical reactions, but with an integrated series of photochemical and ordinary chemical reactions, imposes experimental difficulties which demand the highest order of experimental skill and experience.

In the foregoing pages an attempt has been made to outline various aspects which the phenomenon of plant life presents, and to indicate some of the avenues which have been followed to gain a clearer and more intimate view of plant life and of the conditions which control it. Of necessity, many features have been sketched in barest outline, and some of considerable importance have had to be omitted entirely. The resulting gaps and defects in the picture presented are, there is no doubt, due more to inadequacy of presentation than to lack of material to draw upon. In principal the picture presented constitutes a group effort. Certain aspects have been pursued by the application of particular disciplines, other aspects demand other tools and talents.

Science is essentially an ordered expression of accumulated experience. The tremendous diversity in which plant life manifests itself, and its remarkable capacity for adjustment and modification, make specialization a practical necessity. On the one hand, many of the phenomena of plant life, more particularly those concerned with diversity of form, with natural plant communities, and with the influence of environment on plant development, have been essentially field investigations. On the other hand, the more precise investigations of the composition of natural plant units have demanded controlled experimentation in garden and greenhouse. Finally, reliable experience regarding the various functional activities of plants, entailing exact chemical and physical measurements, must of necessity be secured by means of laboratory methods. Divergent as many of these experiences appear to be at first glance, and although they are often couched in language which is highly specialized, it is becoming constantly more clearly apparent that only by means of a coordination of experience, gained in these different avenues of approach, can a clear and unified expression be attained.

HISTORY

THE SCIENTIFIC BASIS OF THE HISTORY OF SCIENCE

GEORGE SARTON

Division of Historical Research
The Carnegie Institution of Washington

Among the many initiatives taken by the Carnegie Institution for the advancement of research along untrodden paths, one of the least known is the organization of systematic inquiries concerning the past of science. That initiative was taken by the second president, the late Dr. Robert S. Woodward, and continued and expanded by President John C. Merriam. It does credit to both and the more so because it was less appreciated. Indeed even today the number of people, not only in America but abroad, who understand the need of scientific work in that field is amazingly small. Administrators and educators are always ready enough to praise the history of science and emphasize the need of teaching it; they do not seem to understand that in order to write good books on the subject or to teach it well, it is necessary first of all to study it carefully, painstakingly, down to bed rock. If that study is not continued and constantly refreshed by the consideration of new documents or the reconsideration of the available evidence from new points of view, the subject is bound to degenerate into claptrap and falseness.

The matter can be put bluntly as follows: if the history of science is worth knowing (and educators seem to agree on that) then it should be known as well as possible, and any knowledge below the attainable standard should be treated with the same contempt as we would treat half-baked knowledge or quasi-knowledge in any other field, say, geology or American history. If educators think that the critical study of the history of science is too expensive in time and money; very well, let them abandon the subject, for what can be the use of diffusing errors? They cannot have it both ways: if the history of science is as important as they claim it to be, then it should be taken up seriously and not dealt with casually; it should be studied carefully, and no one should be allowed to teach it who had no first hand knowledge of it and had not proved his ability to investigate it, or at least to appreciate critically the fruits of other people's investigations.

Let us suppose you are an astronomer and have been engaged for years in difficult investigations. Maybe you have taken yourself or collected a large number of spectra, and you are making an analytic study of them. When your analysis is carried far enough, you will publish the results in an elaborate memoir, which may (or may not) enrich our general knowledge of astral physics, stellar evolution, and cosmology. Such labor as yours is very difficult, the more so because you have become so fastidious about it, and it would be unbearably dull if you were not sustained by the hope of reaching a new parcel of truth, and thus increasing our knowledge of the subject, however little. What sustains you, above all, is that your inquiries and speculations are not idle, and that they would not be idle even if they proved sterile. You are working with your fellow astronomers in the right direction; you may be out of luck and fail to discover anything of great value; in any case you do not let down your associates, you do not cheapen their knowledge; you hold fast to your scientific ideals.

Almost any scientific investigations which are really worth while imply an infinite amount of drudgery which nothing could compensate and justify, except the thoughts which I have just expressed. In that sense it may be claimed that honest scientific work is one of the highest forms of altruism. The true scientist at the end of fruitless experiments is reconciled to his failure by the fact of his having succeeded in proving their fruitlessness. The path he followed led nowhere, but thanks to his exploration of it other scientists will be saved the trouble of entering it.

To be sure when a laborious technique has to be repeated endlessly, the scientist, however devoted he may be to his task, cannot but experience moments of languor and discouragement. The work is so tedious, the goal so distant, and the reward uncertain. Is it worth while going on? Let us imagine that during one of these inevitable periods of disillusionment you are approached by a friend, "a man of the world," who says to you, "Why do you take such absurd pains to reach an unobtainable aim? Even if you succeed in adding a new chapter, or a new footnote, to astrophysics, what of it? The crown of your labor will be an elaborate and unreadable memoir, which, even if you be lucky enough to have it printed, will hardly be published. For who will read it? Maybe a few other fools like yourself . . . that is all. Instead of wasting your time in that way, why don't you write a book explaining the marvelous facts which we already know about the stars, the numberless universes lost in the incredible depths of space? To think that so many of them have been counted, their prodigious distances measured, their chemical constitution analyzed, their history and fate determined! What a thrilling book one could write about that, a book that might be a best seller, would bring you wealth and fame. . . . Every educated man would know of you; while if you persist in your stupid efforts, the best that you can hope for is the esteem and jealousy of a few colleagues scattered all over the world; a mediocre fame and a mediocre fortune."

This example is imaginary but very plausible; so plausible indeed, and so probable that it requires hardly any imagination to produce it. We may be sure that almost every astronomer has met with that very temptation in one form or another. Some have succumbed to it (and we should not blame them hastily), others have resisted it and continued their task with more or less serenity.

Of course, every scientist will appreciate this kind of temptation and be able to see both sides of it. For it is equally obvious that no scientific progress can be made except by patient and austere investigations, and that the discoveries which are made must be explained from time to time to the public. It is certain that science is not advanced by the composition of readable and salable books, yet such books are much needed for the education of the public, including the scientists themselves, and for the awakening of new scientific vocations.

Far from blaming the *competent* authors of popular scientific books, we should praise them. To be sure if they be truly successful, their financial reward is likely to be much greater than that of pure research and hence they need less encouragement; they received their reward and it should suffice them. I have often thought that elderly scientists should try to popularize their knowledge, and consider that duty as their last one, and not the least. However, only a few scientists will ever be able to do that well, as the performance of that final duty calls for literary qualities which are as rare as the qualities needed for scientific research and somewhat incompatible with them. Even if a scientist had enough literary talent to begin with, the chances are that years of inhibition have atrophied it, and by the time he would like to make use of it in order to fulfill his final duty, it may have died out altogether.

Now the temptation which I have described is far more common in the field of the history of science than in any other field, and far more insidious. Indeed in the other fields the distinction between creative research and popularization is well made and the respective functions and values of each are generally understood. Furthermore incompetent popularization is easily detected and promptly denounced. The historians of science alas! are so few in number, and their field so vast, that they cannot yet enjoy the same protection. While there is perhaps one popularizing physicist for a hundred doing spade work, the proportion is reversed with regard to the history of science. Indeed in spite of all the hullabaloo in favor of those new studies, the number of professional full-time scientists engaged in them is incredibly small. The historians of science who have the privilege of devoting their whole time to their studies and earning thereby a modest living, can be counted on my hands; I mean the historians of the whole world, not of a single country!

On the other hand the number of amateur historians of science, recruited from among scientists, historians, philosophers, and sociologists, is increasing steadily; and the number of popular books, some of them good sellers, if not best sellers, is larger every year. Some of these books are good enough as a first approximation, but most are poor. Nowhere is the backwardness of our studies more obvious. A book entitled "Men of Mathematics" can be successful in spite of its shortcomings, while a book on "Men of Art" or "Men of Religion" or "American Men" on the same level of scholarship would be promptly discarded by a host of competent critics.

This seems to be a vicious circle, for if the competent scholars in a given field are rare, the competent critics are rarer still, and yet criticism is at no time more needed, one would think, than when a discipline is beginning to shape itself. However, the situation is not as desperate as it looks, and the vicious circle will turn out to be a convenient helix, such as symbolizes the method of successive approximations in any field. The helix is still a bit flat, but it will rise gradually higher and higher.

The beginnings are always imperfect, not so much because intrinsic difficulties are not yet overcome, but rather because the difficulties are not yet recognized. As it is very easy for a scientifically trained man with a modicum of literary ability to write on the history of science, the number of imperfect publications on the subject is exceptionally large. It is no exaggeration to say that the majority of so-called historians of science today are blessedly unaware of the difficulties involved, and therefore their writing is very easy indeed, but untrustworthy and to that extent worthless.

To be sure every subject was developed at first by amateurs, but not in the same pernicious manner as the history of science is developed today. The word "amateur" it should be noted has some excellent connotations. An amateur astronomer or naturalist, as there were many in past centuries, is a man who studies natural history or astronomy as a hobby. He does it, not to earn a living, but to spend his life according to his heart's desire. Such an amateur can hardly be distinguished from a professional of today, except that his zeal is if anything more genuine, and that he is ready to take more pains for his pleasure, not less.

The amateur historian of science is more often of an entirely different kidney. He is, for example, a distinguished scientist who has become sufficiently interested in the genesis of his knowledge to wish to investigate it, but has no idea whatsoever of how such investigations should be conducted and is not even aware of his shortcomings. His very success in another domain, the fact that he has long passed the years of apprenticeship, make it difficult, if not almost impossible, for him to master a new technique. He generally lacks the humility of a beginner, and publishes his historical results with blind and fatuous assurance. That is amateurism at its worst.

The good amateurs who were the pioneers of science (including a very few who were pioneers in our own field) were not content to steal the fruits of other people's orchards and let them rot; they were hard workers themselves, working for the joy of discovering the truth. Such as those will still appear from time to time and help us mightily in our advance, but we can never count on them. They may come or not; for the routine work of a developing branch of science, professionals are needed, men having undergone as good a training as possible, honest, and more interested in their studies than in fame or money.

We need not worry so much about the diffusion of knowledge, for that is largely a matter of business. The financial prizes which the successful writers of popular books may hope to obtain are so high that there will never be a shortage of such books. As to their quality, it will depend largely on the general level of knowledge available to specialists and to potential critics.

The specialist's duty is then simply to raise that level. He must continue to produce memoirs which only other specialists will study, but which will make progress in a definite direction possible. The situation is not different for the history of science than for any other scientific investigations.

It is true learned memoirs should not be quite as monotonous or as badly written as they often are. Some aridity is unavoidable, for example, when abundant results are presented in the form of tables or diagrams. One could not conceive anything more arid than a table or a map, which contains no phrases beyond the necessary technical words, numbers, and symbols; yet we are thus offered a synoptic view of complex data, and a better chance of understanding their varied relationships. For the imaginative scientist, a chart is not dull but full of meaning. However, in the writing of the memoirs there is far too much dullness which could be avoided with a little trouble, and should be avoided. The historian of science who is unable to explain the results of his investigations in a pleasant manner is more guilty indeed than other scientists, because we naturally expect from him, on account of literary training, more urbanity.

At this point it is necessary to make a distinction, often overlooked in America, between written and oral teaching. Every writer is a teacher of course, a teacher who never stops teaching as long as his books are read. And just as ordinary teachers may be divided into many grades according to their audiences, from the ones who help the awakening of childish minds in kindergartens, to those who guide a few mature students in university seminars; just so the writers of books teach many audiences, from the lowest "brows" to the highest. By the way, this has nothing to do with the "sale" of their books, for excellent books for the best students may sell very well, while books written as bait for the millions may fail altogether. The written teaching should be as accurate and precise as possible, for the student reading a book has every opportunity of appreciating every word of it. He may read it as slowly as he pleases, referring from time to time to other books, and reread the whole or parts of it as often as necessary.

Oral teaching is essentially different, for an audience, however carefully it may listen, cannot analyze the details, but only obtain a general impression of a subject. The "oral" teacher should have as deep a knowledge of the questions dealt with as possible and be able to give a fluent account of them. Accuracy in every detail can hardly be expected from him and could hardly be appreciated, but his general account should be as close to the truth as conditions permit. In fact it would be wrong for him to overload his account with details which would simply obscure his message without compensation.

The best proof that that distinction is generally overlooked is the common practice of "reading" papers at scientific meetings. It is clear that a paper carefully written for the sake of students who will examine it, each by himself at his own speed, cannot be meant to be read aloud to a group of other men, however attentive the latter may be. To read aloud in public a paper meant to be scrutinized quietly in one's own workshop is just as foolish as it would be to paint dainty miniatures on the surface of large walls. The walls call for broad frescoes; and so do listening audiences wait for general outlines, which they can understand and assimilate at once, not for microscopic analyses which they are unable to follow.

To return to the written papers, which are the true vehicles of expanding knowledge, much confusion is caused by the scientists who insist that historical knowledge need not be as detailed as historians would have it. To be sure a vague knowledge of facts is generally sufficient for all but specialists, and our knowledge of most things is necessarily vague. However, that is not more true for the history of science than for anything else. The situation is the same in every field. We can never be sure that we know a thing well until we know it with the utmost precision down to the smallest details, and it is the tendency of science to increase the precision obtainable, to carry on the search deeper and deeper, and to analyze the data more minutely. Indeed it is in that very way that many important discoveries have been made.

The historian of science has no hope of making great discoveries in little corners, as the scientist does, yet he is always urging himself to pursue the truth just as far as it will take him. After all, he is a kind of detective and judge, and, like the judge, he can never be sure of his facts unless those facts are as closely determined as possible. Vagueness covers a multitude of sins and errors, which can neither be detected nor got rid of, and tend to obscure and nullify whatever of the truth may be mixed with them. On the contrary, the more precise a statement is, the more tangible and correctible it also is. Historians who appear too fastidious or pedantic should not be hastily criticized. It is better after all to go beyond one's duty than to keep short of it. When all the details of a question have been duly analyzed and published, it is not necessary to burden one's self with them in other studies; it may thereafter suffice to outline the question and refer inquisitive readers to the earlier paper for more precise information.

Thus it happens that the basis of scientific history is bibliography, for we are always obliged to refer to various publications either for the original sources or for collateral evidence. Naïve and immature students often dream of avoiding these inconvenient references, by creating a collection of all the sources, or, worse still, anthologies of them. Many collections of texts and documents have indeed been compiled, but no collection has ever superseded completely the previous ones. At best, it has only been "one more collection," and one more perennial reference added to many others. As to anthologies, the idea that any scholar can select the essential, not only for other scholars but even for young students, is preposterous.

Similar fallacies are involved in the frequent requests relative to selections of the "hundred" or the "thousand" best books on the history of science. However well such a selection were made, it would very soon prove inadequate, except for elementary reference. The fundamental difficulty of historical research lies precisely in this, that no limited collection of books and documents is sufficient for every purpose, and that the intense study of any question may lead the investigator beyond the possibilities offered by the richest library or museum.

Amateurish interest in the history of science often takes the form of book collecting. One tries, for example, to gather all the fundamental books of science, the first editions[1] of the great classics, etc.

---

[1] Collectors are always hunting for "firsts." Generally speaking it is not the first edition which a scholar needs most, but the best edition, which may be different from the first and is

No harm in that as long as the public libraries do not spend more than a small percentage of their income for "rarities." Collecting the latter should be left to rich men, because the cost is altogether out of proportion to the usefulness, and has become absolutely indefensible, now that it is possible to "film" any book at a relatively small expense. The microfilm of a book is equivalent to it for every purpose, except fetishism and snobbishness. The great classics are not always indispensable, and they are certainly very insufficient. The richness of a library for our studies depends not only on its possession of the "classics," but also and much more on the possession of thousands or millions of other books, each of which may be important or not in itself, and thousands or millions of scientific and historical papers. It is for that reason that an institute for the history of science owning a large library *ad hoc* would still be very inadequately equipped; it should be located in or near a general library, the richer and larger, the better.

The bibliographic approach to the history of science is fundamental, but it is just that, an approach, and nothing more. The investigation of any question must be preceded by the establishment of its bibliography; yet it is a matter of common experience that the best elements of any bibliography cannot be obtained before the question has been thoroughly investigated. In that sense one might say that the historical study of any question should begin with a good bibliography, and should end with a better one. It is for that reason that bibliographies compiled by mere bibliographers are seldom good; or to put it otherwise, the value of bibliographies tends to increase with the extra-bibliographical knowledge of their compilers. It is true that the mere fact of collecting is a source of knowledge, and no intelligent man can collect scientific books without learning much about the history of science, even if he reads only, in the way of collectors, a few pages of each book; yet there are better ways of learning. It is true also that any good collection is by itself a modest achievement in the right direction; it is like a gathering of the tools for the real workers to play with. The bibliographic and bibliophilic methods should always remain means subordinate to a higher end; should they predominate,

---

likely to include all that is included in the first, plus additions and corrections clearly differentiated from the original text. The determination of the best edition is sometimes difficult; different editions may be each the best for different purposes.

scholarship would be debased. The bibliophilic approach is perhaps the most amusing of all; it is at one and the same time the most expensive and the poorest.

Scientists have a tendency to believe that historical work is too easy and they despise it accordingly, considering it as a kind of *refugium peccatorum* for their weaker brethren. On the other hand, they are often irritated by the "superior" attitude of some historians, not the best, who survey experimental efforts and the infinite drudgery of research, as it were, from an Olympian height. Historians, whose own scientific education is imperfect and too artificial, who never did genuine experimental work themselves, are apt to behave in the very objectionable way of those philosophers or theologians who were so confident in the superiority of their point of view that they did not hesitate to judge the efforts of scientists without taking the trouble of making a serious study of them. For example, remember the nonsense written by the Scotch metaphysician Sir William Hamilton (the "other" Hamilton) against mathematics, or the remarks on Darwin's theories made by the Bishop of Oxford at the Oxford meeting of the British Association in 1860.

Historians whose judgments would be as reckless as these would be equally contemptible, but it is well to bear in mind that even when historians are as honest and painstaking as can be, they are still obliged to judge other people's efforts in a manner which must seem to the latter summary, abrupt, nay, offensive. It is in that sense that we often repeat the old saw, "criticism is easy but art is difficult." Criticism, good criticism at least, is not easy, but as compared with the criticized achievement it is an extremely short cut. Think of an explorer trying to reach the magnetic pole, or the top of the Himalayas, think of his courage and tenacity, of his sufferings, of his endless devotion; and then visualize the critic, who we shall assume is a competent and honest one, sitting in a comfortable armchair and reviewing, analyzing, and finally judging the other man's sacrifice within a week or less. Or think of a physicist repeating with equal faithfulness and ingenuity thousands of similar experiments, and then again visualize his critic, a competent and honest one, reviewing and judging that immense effort in a few pages or perhaps in a few paragraphs!

One can sympathize with the physicist's or the explorer's impatience, even when it is not justified. For the crux of the situation is this, that the creator is generally not the best judge of his achievement; he is too close to it, too keenly aware of the technical difficulties which he has overcome, and very often, almost unavoidably, too narrow in his own outlook. Now, to take a single instance, the value of an achievement is largely independent of the technical difficulties involved, and it is very hard for a technician to recognize that. The fact of overcoming great difficulties is praiseworthy but only to the extent that their overcoming was sufficiently worth while. Some clever technicians evoke in my mind the image of a man crossing the Niagara Falls walking on a tight rope. That is very difficult indeed and extremely dangerous, yet the achievement leaves me cold; I would hardly dare say so to the courageous artist who did it for I would be afraid of hurting his feelings. The scientific critic is not obliged to imitate my discretion; or rather his duty is to speak out and often to judge summarily scientific achievements which are as arduous and withal as futile as crossing the Niagara on a tight rope.

The paradox involved in the conflict between creation and criticism, or creator and critic, is at the bottom of all the misunderstandings between the scientist and the historian. I have discussed it before,[2] but must solve it again for the sake of rounding off my argument. Who would dare say that I have no right to criticize a dish because I am incapable of cooking anything myself? Creative efforts require a number of qualities which are very different from the purely critical ones, and sometimes antagonistic to them. It is not good for a creator, whether he be a scientist, or an artist, or what is commonly called "a man of action," to be too critical, for criticism would inhibit his efforts, and might even paralyze him altogether and dry up the source of his inspiration. This was neatly expressed by the French composer Rameau when he realized how his creative power waned in proportion as his critical power waxed: "From day to day my taste improves but I have lost all my genius."[3]

---

[2] See my preface to vol. 29 of *Isis:* "Primum herbam, deinde spicam, deinde frumentum plenum in spica" (*Isis*, vol. 29, pp. 9–14, 1938).

[3] As quoted by Catherine Drinker Bowen and Barbara von Meck in their beautiful biography of Petr Ilyich Chaikovskii, "Beloved Friend" (p. 15, New York, 1937).

It is the historian's first duty to consider the facts of the case very carefully before formulating a judgment. He must make sure by all available means that the facts considered by him are pure and eventually disentangle them from the dross of errors and misunderstandings surrounding them. Thus far, he is doing what every scientist does, but there is a fundamental difference. The other scientists, except the anthropologists, have only material facts to take into account, and such facts, however complicated, are relatively simple as compared with human facts. Now the historian of science must be acquainted not only with a large number of material facts, but also with human facts. His every judgment implies some human facts. When he says "Argon was discovered by Lord Rayleigh and Sir William Ramsay in 1895," he expresses the conclusion of an analysis involving not only the physical and chemical conditions leading to that discovery, but also the personal reactions of two remarkable individuals. Indeed the main difficulty lies there, and an equitable determination of the respective merits of both discoverers is a very subtle matter requiring not only technical knowledge, but wisdom.[4]

This example is not exceptional but typical. Every scientific achievement is primarily a human achievement. The historian of science is not a historian of nature (or "naturalist") but a historian of man. He is in the highest sense an anthropologist. It is rather intriguing, in that respect, that while the common anthropologist has long been accepted into the scientific fraternity, the historian of science is still held out of it. Thus the study of primitive thought is considered a legitimate scientific subject, while the study of the thought of the great pioneers and intellectual leaders of mankind is pooh-poohed! This is not quite as absurb as it seems. The anthropologist has become a true scientist because he has undergone a true scientific training, and because he has finally succeeded in driving the dilettanti out of his domain. The historian of science is still necessarily an amateur, and, more often than not, a bad one. As historical dilettanti speak louder and to larger audiences than the austere investigators, they seem to dominate the field. In short, the historian of science has not yet won his claim to respect-ability. He is still an outsider and an outcast. Is it not high time to give him a chance of redemption? Educators who praise the history of science to the skies but can find no room on their staffs for historians of science have a share of responsibility in that pitiful situation.

---

[4] For a discussion of that particular case, see *Isis,* vol. 8, pp. 178–179.

The human elements, with which the historian of science is constantly dealing, increase the difficulties of his task immeasurably; indeed they increase it to the danger point, for the analysis involved is often so subtle as to be beyond the discipline of scientific method. On the other hand, they give to his studies their deep humanistic value. It is because of them that the historian of science is a kind of mediator between the scientists and the humanists, and is able (or would be able if given the opportunity and the means) to harmonize and unify whatever culture we may possess. This has been recognized, almost instinctively, by a number of distinguished scientists and scholars, but educators have not yet understood it, or at any rate have not yet translated their understanding into deeds.

The sayings of scientists on this subject are generally better known than those of scholars, and therefore I take pleasure in quoting here a statement made as early as 1880 not by a scientist proper but by a great scholar, the Sanskritist Max Müller:[5]

> You know that I am not a man of science in the usual, though far too narrow, sense of the word. But, although the science to which I have devoted the whole of my life, the science of man, is not yet formally represented in this College, depend upon it it will enter in; it is there already, it lurks in every corner, and I trust it will soon fill the whole College with its genial warmth and its quickening impulses. You mean to teach mathematics. Can you teach mathematics without teaching the laws of thought, without telling your pupils something about such men as Thales, Pythagoras, and Euclid, who were ancient Greeks, but who were men of science for all that? You mean to teach physiology and biology, the laws of life and nature. Are you likely to leave out the very crown of nature, man—to leave nature, like Hamlet without the Ghost, a nature without its spirit? A true College of Science could not live if it were to exclude the science of man. Man is the measurer of all things, and what is science but the reflection of the outer world on the mirror of the mind, growing more perfect, more orderly, more definite, more

---

[5] He made it in the course of a speech at the opening of the Mason Science College, Birmingham, in October 1880. The quotation is taken from "The Life and Letters of the Right Honourable Friedrich Max Müller," by his wife (vol. 2, p. 89, 1902).

great with every generation? To attempt to study nature without studying man is as impossible as to study light without studying the eye.

Max Müller had given but little thought to the history of science conceived as an independent discipline (hardly anybody had in those days), but he understood deeply the necessary interpenetration of the sciences of nature and the science of man. Every historian, and more particularly every historian of science, is obliged to appreciate human achievements, which is difficult, and to judge the men themselves, which is far more difficult. Yet that duty is unavoidable. One of the functions of history is to explain men's motives and conduct, to criticize their deeds, to take warning from their errors, and to take advantage, as much as we can, from their virtues. The historian of science must explain the conditions of scientific discoveries and the evolution of theories; he must also introduce to the reader the men of science themselves, as men of flesh and blood, and tell the vicissitudes of their lives as truthfully as possible. Great scientific achievements are rare, great men of science are rarer still. It is important to describe their greatness when it occurs, and thus create noble emulations. The lives of such men as Faraday, Darwin, Pierre and Marie Curie deserve to be studied as the lives of saints, saints of a new kind, new models for the youth of today. It is the historian of science's duty to draw the portraits of those great leaders, and to hold them up for remembrance, gratitude, and devotion. In this way he continues and fulfills the old tradition of history as a mirror (*speculum*). It is his business to polish the mirror and to watch it as jealously as the astronomer watches his own mirrors. These reflect the stars, the historian's mirror reflects man. What is more important? I am not going to tell you, but please consider that the astronomer's mirror reflects the stars, and that the historian's mirror reflects the astronomer himself!

The history of science is a part of the humanities, an essential part of them. It is a science, because its purpose is to classify and organize a definite group of objective facts—but like every science, or more so, it is also an art. One can never consider all the facts, and even less classify and organize them. One must select one's data even as every artist selects his. The indetermination extends into two infinite directions, for in the first place, the facts are largely human facts and hence not as clear cut as material facts but highly capricious and often evanescent; and in the second place, one must choose a few of the achievements among a great many, and a few men of science among thousands. To hold up to one's fellow men the mirror of history is a sacred function, the fulfillment of which is as difficult as anything can be. It is high time that efforts should be made to polish a good mirror for the history of science, for as long as that is not done, the mirror of history in general, the great mirror, will be very imperfect and very incomplete. What the historian of science is called on to do is to reflect carefully a group of facts which other historians have unaccountably neglected; these facts are not casual and irrelevant, but on the contrary highly relevant, yea, essential for the understanding of human progress and destiny.

It is not easy to find good men for that task, because there is not as yet an institute where they could be systematically trained and where every-day cooperation with their fellow workers would stimulate and enlighten them, and also because the needed qualifications are very complex. The minimum requirements are: experimental knowledge of science in general, and deeper knowledge of one or two branches of it; sufficient polyglottism; historical and philosophical training. In addition, there are other requirements which are less tangible but hardly less important. For example, there are ways of testing a man's acquaintance with historical facts and methods, but it is far more difficult to determine whether he has any historical sense. The same remark applies to philosophical knowledge, and indeed to every kind of knowledge. A man's actual knowledge is far less important than his ability to obtain further knowledge and to use it in the best manner at the proper time.

The historian's value increases with his experience and wisdom. These two terms are frequently opposed as if the one represented something which can be gradually acquired while the other is an outright gift. That distinction is not as correct as it first appears to be. To begin with, one's wisdom may perhaps be improved, at least within a certain (small) proportion. Every human quality can be stretched out somewhat if sufficient pains are regularly taken to that effect. On the other hand, our experience will grow only if we have the gift of acquiring and retaining it, and of storing it properly in our brains.

What is clear is that the main qualifications of a historian are innate gifts, comparable to the gifts which make possible the development of a good scientist or a good artist. The historian of science, like any good specialist, must have a vocation; and if he have it, he must be prepared (but this is but another part of his vocation!) to take infinite pains in order to improve his qualities and make the best of them. The task to which he is addressing himself is so large than he cannot expect to succeed unless he devotes himself to it whole-heartedly, and not for a few years only, but for a lifetime.

The historian of science is often a scientist who is more interested in ideas than in technicalities, and who may have grown tired of laboratory work. It is a great mistake, however, to imagine, as is sometimes done, that failure to do well in the laboratory is a promise of success in the library. The would-be historian making that mistake is like the farmer's son who fancied he had a call to the ministry, because work in the fields was so hard and hot. These things have no causal relationship. Indeed if the failure were due to a lack of intelligence or application, it is clear that it would be repeated; if it were simply due to a lack of manual dexterity or instrumental ingenuity, it might be avoided in the new domain. Vocations, it is well to remember, should be characterized by positive qualifications, not by negative ones.

I was approached a year ago by the representative of a great institution, which desired to organize the teaching of the history of science. They were willing to pay the right price to the right man, but they expected him to give a spectacular course, which would have impressed vigorously as many students as possible. They thus betrayed a misunderstanding of the true function of the history of science. It is clear that it would be hardly possible to impress a large body of students, already inured to high-pressure advertising in all its forms, without resorting to the cheap tricks of popular authors or popular lecturers, without overemphasizing the points appealing to their uneducated fancies and neglecting correspondingly points of lesser attractiveness but of greater significance. No, the teaching of the history of science should be as austere as the study of it, austere and humble. This does not mean that it should be unpleasant and ill shaped; on the contrary it should be as elegant and urbane as possible, but without any use of "purple patches" and without any appeal to attractive but irrelevant notions.

The first need is to study the history of science as honestly and humbly as the biologist studies protoplasmic reactions or the physicist investigates the properties of matter. The second need, which can only be fulfilled in proportion as the first is, is the teaching of the truth thus discovered, in the same spirit, that is, the teaching of it in a simple, honest, and quiet way; as clearly and beautifully as possible but without extravagant claims or dramatic gestures.

There is an immense amount of work to be done with regard to the discovery of the truth as well as for its diffusion; it is a great and far-reaching undertaking for which the best men will hardly be good enough. Even from those men one should not expect too much, or at least anything sensational. We would expect them to do their best, to carry on honest investigations like other scientists and scholars worth their salt. Once in a while, there may be a genius among them, one seeing much farther than his fellows and making a new synthesis, but that is a gift of the gods which can neither be foreseen nor bargained for.

THE SUBSTANCE OF AMERICAN HISTORY

WALDO G. LELAND

Permanent Secretary, American Council of Learned Societies
Research Associate, The Carnegie Institution of Washington

The history of the American world offers opportunities for research of unsurpassed interest and importance. A vast area of the most varied character, of great natural resources, thinly populated by native stocks possessed of widely varying degrees of culture, was opened to European penetration at the close of the fifteenth century. Thereupon commenced a process that has continued to the present day, a process characterized by geographical exploration and the progressive occupation of new territory, by the conquest of native populations by European invaders, by conflicts among the invaders themselves for possession and control, by movements of population on a scale previously unknown, by the regrouping of European stocks in a new environment and in new political, economic, and social combinations, and by a long series of experiments covering the entire range of human activity.

The effect of this process upon the history of the entire world has been incalculable and has steadily increased during more than four centuries. The nature of the process, its conditioning factors, its details, the interplay of the forces involved, and the succession of phenomena emanating from it constitute the substance of American history.

The permanent factors of history have been variously estimated at different times; sometimes historians have been inclined to give weight to one, sometimes to another as being preponderant. Thus geographical and climatic influences, race, economic forces, politics, ideas, and religion have had, each in its time, the major share of the historian's attention as from age to age the interest of mankind has shifted from one point to another. It is very nearly true that each generation must rewrite the history of the past in order to bring it within its own frame of reference. For history is not merely the objective ascertainment of facts, which can be observed and determined once for all and thereafter accepted without question. The ascertainment of facts, by methods as nearly scientific as possible, by which, of course, we mean those methods that seem most likely to assure reliable results, is indeed the ostensible purpose of the historian, but in spite of himself and in spite of a genuine desire to be completely objective, the necessity of interpretation brings into the labor of historical investigation a subjective factor that cannot be suppressed.

This factor expresses itself in the selection of the facts believed to be most pertinent and illuminating, and in the arrangement and presentation of the data yielded by the investigation. The influence of the subjective factor is heightened by the fact that the historian is dealing with matters of human conduct with respect to which he has convictions, ideas, or notions that are so much a part of his mental makeup that he cannot divest himself of them. Even if he is able—and often he is able, especially if he is dealing with subjects that do not cause strong emotional reactions—to assume an attitude of complete impartiality, nevertheless the manner in which he balances and judges his evidence is determined by his intellectual personality. From these limitations no human being is exempt. The astronomer or physicist or mathematician, engaged in similar inquiries, would be subject to the same influences, and perhaps to a greater degree than the historian, who has trained himself to resist them.

The conclusion to be drawn from these conditions of historical research is that history is not and cannot become a science, in the sense that physics and chemistry and geology and astronomy are sciences, nor is the cause of historical research benefited by claiming for it a scientific character that it cannot possess. At the best we may hope, by means of diligent investigations, conducted according to the various skills and techniques that we call historical method, and in a state of mind appropriate to scientific inquiry, to ascertain the approximate truth with respect to human conduct in the past.

The nature of our knowledge of the past has a determining influence upon our attitude toward the present and upon our conception of the future. We can no more escape from this form of determinism than the sane individual can escape from shaping his conduct in the light of his experience. He may, indeed, appear to disregard his experience, but even that is a negative manner of determining his present line of action. For this reason the historian is especially subject to pressure of all sorts from without. Even the power of government is sometimes employed to make him present the past in such a way as to justify certain policies and modes of action. Economic and social interests are also eager to have his support, and offer inducements or make threats in order to secure it. Groups of well-intentioned persons who desire to promote the common good in accordance with their own views as to the nature of that good combine to procure the historical rationalization of their doctrines. To influences of these various sorts the historian is peculiarly subjected, and against such influences he must develop uncommon powers of resistance.

With this understanding of the limitations of historical research, and of the dangers to which it is exposed, let us turn to a closer contemplation of the history of America and of the major factors in the historical process that was set in motion by the European invasion. And, although the process in question has operated throughout the entire American world, imparting to its history a distinct character of unity, it is permissible to confine our attention particularly to the history of that part of the area comprised within the United States, as being representative of the process as a whole and as illustrating with especial vividness certain important aspects of it.

In any inquiry into the history of a given area, it is essential that we should understand the nature and influence of the physical environment of the people whose conduct we are investigating, their own physical and spiritual characteristics, and their relations among themselves and with other groups.

The people of the United States have been influenced by their physical environment in countless ways. To the Europeans who first penetrated into that area—the Spaniards, French, English, Dutch, and Swedes—it was chiefly a country of vast forests, thinly populated by native stocks in relatively low states of culture, whose conquest, although attended by nearly constant conflict over a period of two centuries, hardly interrupted the course of the invasion.

The disposition of rivers and mountains exercised an important influence upon the relative fortunes of French and English colonization. The English colonies, situated near the center of the Atlantic seaboard, were backed by the ranges and forests of the Appalachian system, which confined them to a narrow area, forcing them to a compact growth and to the development of agriculture, industry, and sea-borne commerce. The French, on the other hand, entered the continent by its two great waterways, the St. Lawrence and the Mississippi, and these avenues into the interior led to the dispersion of the colonists and to their dependence upon trade in forest commodities, chiefly furs, with the Indian tribes, with the result that the development of agriculture and industry was retarded. These influences, which favored a rapid increase in the population of the English colonies while that of the French colonies increased much more slowly, gave to the former an overwhelming advantage in the final struggle for the control of North America.

Geographical influences do not, however, remain constant, for they themselves are modified by man. The forest areas of the United States have been drastically reduced; the mountains that confined the English colonies have yielded up their stores of ore, coal, and oil, and have become centers of heavy industry; the great rivers of the continent have been superseded as lines of communication by railroads, and these have in turn been supplemented, and in some cases supplanted, by the rapidly growing system of highways that the universal use of individual automotive transportation has made necessary. The water power of New England, which determined the location of the first industrial centers, has lost its significance as other forms of power, and especially of distributable power, have been developed. The natural grazing grounds of the continent have been transformed into wheat lands in wet periods and into arid areas of dust in dry periods; former desert regions have yielded to agriculture through dry farming or irrigation, and soil erosion has made vast areas unfit for cultivation. River bottoms and swampy lands once infested by malaria have been cleared of malaria carriers and made fit for habitation. To such an extent have its inhabitants been able to modify the surface of the United States that, although a map will show the same major features as in the seventeenth century, their significance has been greatly changed.

Other physical influences remain more nearly constant, although their effects have been modified. The regions of geological instability are still subject to earthquakes; tornadoes still cut narrow paths through the interior; hurricanes still ravage the coasts of the Gulf of Mexico and the Atlantic; and flood waters still descend the river valleys. Despite periodical variations the climate of the United States has not changed in four centuries. The problem of the historian, which is to ascertain the effect of such physical influences upon population and upon human conduct, makes it necessary for him to turn to other disciplines, such as geography, geology, meteorology, and medicine, to mention only some of the more obviously useful ones, for their cooperation.

In such a physical environment, a population characterized by constant changes in its composition has grown up. It has always been a population of mixed stocks, but the mixture has undergone many transformations, and the nature of these constitutes one of the most important problems in American history.

In 1790 the white population within the present area of the United States numbered some three and a quarter millions; of this number, three-fourths were English with small proportions of Welsh, Scotch, and Irish, and the remaining fourth was made up of German, Dutch, French, Swedish, and Spanish elements. To this total must be added, of course, a large population of Negroes of African descent and of mulattoes of mixed strain, as well as of Indians, but the latter were external to the rest of the population rather than an integral part of it, as they became in Mexico and certain other southern countries.

In the nineteenth century there began a movement of population into the United States on an increasing scale, which, originating in the British Isles, in Germany, and in northern Europe, came, toward the end of the century, chiefly from southern and eastern Europe. The result of this movement was a mingling of cultures and a mixture of stocks which is still in progress, and the outcome of which can as yet only be forecast with uncertainty. But the process provides tasks of fundamental importance for the historian, the anthropologist, and the sociologist in reconstituting the cultures of the immigrant groups and in estimating the influence upon them of their new environment and of their impacts upon each other, as well as the effect upon the whole mass of the population of their incorporation into it.

This study of cultural content and change calls for the cooperation of many disciplines. Such basic elements as language, musical and artistic expression, literary production and taste, folklore, handicrafts, etc., must be recovered, described, and evaluated, and the techniques of ethnology, linguistics, musicology, literary history, and history of art must be applied to the different aspects of the general problem.

The history of ideas and of opinion must also be investigated. To what extent can "American" ideas, as distinguished from "European" or other ideas, be shown to have developed and taken root during the generations of colonial existence, and how were these affected by the ideas brought in by the immigrant groups, and how were the latter affected by the "American" ideas that they encountered? The field of American religious history is also a fertile one, for the intellectual climate of the United States has been favorable to the rise of religious and sectarian movements, some of which have spread into other parts of the globe. Closely akin to the history of religion is the history of morality, or rather of ideas and opinions respecting moral conduct, whether or not supported by religious sanctions. Sometimes known as moral standards, such ideas have varied widely from time to time, from place to place, and from group to group; always they have exercised a very considerable influence, and the history of their effect in stimulating or restraining conduct, and in promoting legislative efforts to enforce certain standards or practices, is of major importance.

The relation of geographical influence to cultural development constitutes a phase of American history that is worthy of special attention. It is well known that the isolated groups of the southern highlands have preserved almost intact many elements of culture that they possessed in the eighteenth century and that are survivals of still earlier English forms. The cultures of the plains differ from those of the seaboard, while cultures of a still different sort have developed along the great river highways. But cultures are constantly being modified, and the acceleration of that process as new communications are established between regions and groups has become a phenomenon of much significance.

The intellectual history of the United States affords another approach to an understanding of its population. The progressive advancement of knowledge in the various fields of science and learning, the history of research and of the far-reaching consequences of the application of its results in medicine, industry, and engineering, as well as in the lives of individuals, are subjects of inquiry that concern the specialists in the fields of science and technology, and their cooperation is essential to the historian. An active interest in education and especially in making it available at all levels to the largest possible number has been characteristic of American thought, and the history of educational organization and philosophy in America is of particular importance for an understanding of American history.

Finally, it is essential to know as much as possible of the physical changes that have taken place in the American people. Are the individuals of the present generation different physically from those of earlier periods? Are they taller; do they live longer; to what diseases are they most subject as compared with those of earlier generations; what variations have taken place in the fertility of the various stocks; what physical evidences are there as to the course of mixtures of stocks during the past century; and what have been the results of such mixtures? These and questions of similar import are fundamental in any attempt to explore the history of the population of the United States, and the historian must again have recourse to other disciplines, such as biology, anthropometry, medicine, and statistics.

American historians have chiefly concerned themselves with what is commonly known as political history, but with increasing attention, in late years, to economic and social factors. It is of course impossible to draw sharp lines of demarcation between the political, economic, and social conduct of human beings living in close association, although it is convenient, and even necessary for purposes of study, to distinguish between different kinds of activities, and this necessity gives rise to the special disciplines of political science, economics, and sociology.

All American life has been profoundly influenced by the prospect of a new world possessed of apparently limitless natural resources waiting to be exploited, and this conception of opportunity has made Americans desirous of retaining the utmost liberty of action for the individual and has also made them impatient of social control. This attitude, reinforced by the tradition of personal freedom that grew out of circumstances of colonization, and that was further strengthened by the war for independence, has marked popular political and economic thought during most of the history of the United States. But with the exhaustion of free land, the disappearance of the frontier, an increasing shift from agriculture to industry, and the concentration of population in urban centers, a willingness and even a demand for increasing social control have manifested themselves.

Along with this transition has come a corresponding change in attitude toward the centralization of governmental powers. For the English colonies it was a novel idea to be represented in occasional conferences of delegates to consider matters of common interest. Even the war for independence produced only a loose confederation, hardly more than an alliance, of independent sovereign states. When a closer federal union was formed, its status and stability had to undergo the test of civil war, but with the assumption of imperial responsibilities after the Spanish War, the position of the federal government rapidly became more dominant, until participation in the World War and the difficulties of the prolonged depression caused it to overshadow altogether all forms of local government, and to extend its activities into every area of American life.

This trend toward centralization and extension of governmental control has been accompanied by corresponding changes in economic and social views, and by demands for increasing regulation of economic interests. Changes in industrial and agricultural organization in the direction of large-scale operation and mass production have tended to emphasize and perpetuate differences in status between capital and labor, with resulting conflicts, in which labor, whose political influence increases, demands the intervention of government in its behalf. During this process the attitude toward property also has undergone change and the rights connected with its possession have tended to be extended to the opportunity to gain a livelihood.

Social life has also undergone changes of similar magnitude. Organized religion no longer furnishes the chief centers of social activity as it did until toward the close of the nineteenth century. With the universal possession of individual means of rapid transportation the isolation of communities and of rural homes has ended, and the cinema and the radio have brought to every corner of the land the same diversions, the same ideas, the same employment of leisure time. Social life has become nearly uniform and has, so to speak, been nationalized by the results of scientific research and invention. The rapidity with which this change has taken place, within less than half a century, makes it all the more striking and also makes possible intensive study of it by the historian, the sociologist, and the economist on the basis of direct evidence.

The attitude of the people of the United States toward other countries, upon which must be based the foreign policy of the government, has remained more nearly constant during a century and a half than opinions respecting many of the matters that have just been referred to. A geographical influence of the greatest importance is the situation of the United States between two great oceans. which has favored a sense of isolation and of remoteness from the countries of Europe and of Asia. In addition to this geographical factor is the fact of migration from Europe to America and the accompanying feeling of having definitely left behind the affairs and interests of the Old World. To offset this feeling, there are, it is true, sentimental ties that unite emigrant groups in varying degrees to their respective countries of origin, and as a result of these ties there are also to be found among the same groups a reflection of national antipathies. However, in spite of the feeling of remoteness referred to above, it is certain that interest in the affairs of European and Asiatic countries has greatly increased during the twentieth century, and participation in the World War, in spite of reactions that followed it, has left the conviction on the part of increasing numbers that the United States cannot be indifferent to whatever affects the rest of the world or any large part of it. At the same time, a sense of defensive solidarity with the other American countries has emerged, and has been accompanied by increasing interest in them and growing appreciation of their cultures.

The effort has been made to indicate a few only of the most important elements in American history and to distinguish a few general trends. The result is of necessity a very incomplete and much oversimplified picture, outlined with broad strokes, but it is hoped that some idea has been presented of the substance of the history of the United States and of its bearing upon our efforts to achieve a better understanding of the nature of mankind and of human destinies.

No single program of research can be as broad as the whole field, but any program should have a clearly defined relation to significant problems. Where such relation is lacking, history degenerates into antiquarianism. It must also be clear that the historian is forced to seek the cooperation of many other disciplines, some of them highly specialized or technical, if he is to gather and understand all essential evidence. Thus viewed, most historical investigations become cooperative undertakings, and their results have significance for a wide variety of intellectual and social interests.

THE CORRELATION OF MAYA AND CHRISTIAN CHRONOLOGIES

A. V. KIDDER

Chairman, Division of Historical Research

J. E. THOMPSON

Division of Historical Research
The Carnegie Institution of Washington

No longer, fortunately, does research need the justification of promising immediate returns of material value. But even this happy condition has its dangers. Removal of pressure for "results" can, and often does, lead to haphazard undertaking of projects without due consideration of their setting in the general scheme of knowledge. Consequent failure to visualize relationships tends to prevent that clear perception of ends to be attained which is so necessary for the precise formulation of problems, and for intelligent and straightforward attack upon them. Particularly is this true of disciplines dealing with the career of man, a subject so vast, so baffling, and which ramifies so fascinatingly that effective hewing to the line is extraordinarily difficult. For this reason it is most essential, in organization of study in any one of these disciplines, early to visualize major objectives; and, as work progresses, to assure oneself that effort is being directed toward answering questions of real significance.

The above platitudes are offered in excuse for the presentation of certain other equally self-evident, but sometimes not sufficiently considered, facts.

Archæology and history seek to picture man's past; ethnology and the other social sciences to record the conditions under which he lives today. Both groups gather and analyze data necessary for understanding the nature of man and the structure of the modern world. And all of them have to deal, in one way or another, with that characteristic possession of the human race, that extra-organic phenomenon which anthropologists call culture, and which in its more highly developed form is known as civilization.

Every animal save man comes into the world endowed with all, or very nearly all, the knowledge needed to carry on its individual existence. Man, on the other hand, is born intellectually naked. Nothing is transmitted to him from former generations beyond his physical makeup, a receptive but empty mind, and the urge to exercise such purely organic functions as feeding and reproducing himself. All else is acquired in postuterine existence by the teaching-learning process. And while it is true that some of the so-called higher animals are capable of making certain limited additions to their chromosomal stock of knowledge, and although the evolution of the human brain has doubtless rendered that organ increasingly susceptible of receiving and profiting by instruction, the fact remains that man has set himself widely apart from his fellow creatures by surrendering to his culture almost his entire equipment of innate abilities.

Culture, being extra-organic, is unrestrained by the slow-working laws of biological evolution. Also, as culture is the product of many and exceptionally gifted individuals, it constantly tends to outgrow the ability of the average man or, indeed, the capacity of any single individual to comprehend it in its entirety, or to make use of it in the most effective manner. Culture therefore can, and history teaches us that it repeatedly has, become so overwhelmingly complex as to bring confusion and ruin to those very groups which have been most active in its creation. Thus, although it is man's greatest triumph, it is also his most dangerous possession. Of this truth, it is perhaps unnecessary to state, the most recent and most striking manifestation is offered by the perilous stresses now being brought about by current advances in the physical sciences.

We must, then, if we are not to risk eventual destruction by this Frankenstein's monster which we have created, know far more than we do today regarding culture: its origin, its growth; and the laws, if such there be, which govern its reactions both upon individuals and upon racial groups.

Of the social sciences, anthropology has made the study of culture most peculiarly its own. And to archæology, that branch of anthropology which deals with prehistoric peoples, is delegated the task of recovering as much as possible of the story of the rise and spread of early cultures and of their flowering into the various historic civilizations.

Archæology, of course, is limited not only to consideration of the material aspects of culture, but also to those of its manifestations which have produced imperishable remains. Its field is enormous, both in space and in time; the evidence at its command is pitifully fragmentary. But a great deal has already been learned. We are aware, in rough outline, of the succession of the Stone Ages in the Old World. We know that man has been a tool-using animal for something like a million years, during the earlier part of which time, we must suppose, there took place the mysterious exchange of inherited instinct for the ability to learn and to profit by individual experience. We also know that the growth of culture was so exceedingly slow that not until the very verge of the present, not, in terms of the whole period, until yesterday, did there begin that swift acceleration which has rushed us upward through the Neolithic, the Ages of Metal, and into the rocket-like surge of modern civilization.

What is the nature of this terrible vehicle to which so unreservedly we have intrusted ourselves? Are there, as has already been asked, laws which control it? Answers to these vital questions can only be possible when we have far fuller information as to the history of culture as a whole; and, in particular, regarding the rise of the higher civilizations.

Where do we stand at present? We know that the civilization which we have inherited arose in the countries bordering the eastern Mediterranean; that another grew up in the Far East; and that a third essentially similar, though seemingly later, development took place in the pre-Columbian New World. We do not, however, know whether these three civilizations were genetically related, or whether each was an independent manifestation of some innate urge which, given proper environmental conditions, impels all human beings to take certain cultural steps.

Solution of this obviously key problem cannot be reached until we are in possession of detailed information as to the material content and the relative chronology of the Mediterranean, Far Eastern, and New World civilizations. It is the duty of Americanists to collect, in the Western Hemisphere, their share of the data necessary for the worldwide comparative study which will ultimately permit sound conclusions.

The Indians were not granted opportunity to demonstrate whether or not they were capable of reaching the levels of advancement attained in Eurasia, because their progress was abruptly and conclusively halted by the advent of the white man. This, from the point of view of the anthropologist, is one of the great tragedies of history, for even if it should be found that the seeds of the higher New World cultures came from Asia, it is certain that all significant communication between the two hemispheres had been cut off for many centuries before Columbus. Thus the Indians were working out their own salvation, and they seem to have been traveling a road remarkably closely parallel to that followed by Old World man, achieving much the same triumphs and making much the same mistakes.

It is for this reason that the Americanist regrets the Discovery. He would so greatly like to know what would have happened had the Indians enjoyed another thousand years of isolation. Would they have continued to develop along Old World lines? Would, here as there, the torch of progress have been passed from people to people with the rise and fall of nations? Or would there have come into being quite new types of human society? One can perceive foreshadowings of novel experiments in communal life on a large scale. But the white man burst in with fire and sword, the flimsy, immature political structures of the Indian fell instantly to ruin, and his upward struggle came to naught. The loss to mankind may have been great.

We should, however, be thankful that the Conquest did not occur in the days of the Romans or, worse still, at the time of Alexander. For the greater part of New World progress apparently took place after 500 B.C., and an earlier overrunning of the Americas would have left us with far less information than we now have, or rather than we can have, for our much-needed comparative study of civilizations.

The qualification just expressed is of course due to the fact that we do not as yet know the hundredth part of what further research can tell us about the pre-Columbian career of the Indians. The ground has been no more than scratched. Assiduous and intelligent archæological work can be counted upon to build up a very respectable history of the New World cultures, determining whether or not they were autochthonous; revealing, if the latter, their point or points of origin, their spread, their interaction, and the relative ages of their various regional manifestations. And very special effort must be made to gauge the time factor, for in the absence of actual dates we would remain ignorant of the chronological relation between Old and New World developments, knowledge of which is essential for reaching a decision as to the degree of their relationship.

Our task, then, is not only to reconstruct New World prehistory, but also to date it. The latter, however, we shall probably never be able to do as accurately as has been possible in the Old World, because, although certain groups of of Indians rivaled in many other ways the progress made in the eastern Mediterranean and the Far East, none of them, with the single exception of the Maya, had reached the point of making permanent record of the lapse of time. Lacking practically all documentation, we are forced to carry out our chronological attack by methods less precise. These include various types of estimation: by depth of deposits, the state of preservation of artifacts and buildings, growth of trees, changes in culture, various geological phenomena. For help in such studies the archæologist must call upon specialists in numerous other disciplines. But for nearly all of them, and in particular for the gauging of time by developments in the arts, it is essential to have starting points in the form of known sequences of events which can be rated chronologically.

Few sequences of this sort are as yet available to us. The lists of Inca and Aztec and Toltec rulers set down after the Conquest yield a certain amount of information, but like all similar records they tend to become unreliable the farther back they go. It therefore behooves us to make the fullest possible use of the two exact chronologies which we have.

The first of these is based on A. E. Douglass' brilliant research upon the growth rings of trees used as beams in the construction of prehistoric dwellings. This has resulted in the absolute dating of more than a thousand years of culture history in southwestern United States.[1] The dates recovered by Douglass and his pupils have repeatedly been checked and cross-checked by competent Southwestern archæologists, and while opinions may differ as to the climatological, astronomical, or botanical significance of the tree rings, their reliability as time markers is beyond question.

---

[1] For the clearest exposition of the dendrochronological method, see Douglass, 1935, a paper which also contains a comprehensive bibliography.

The second absolute New World chronology is that of the Maya. The culture of this remarkable Central American people is of undetermined origin, but it is thought to have stemmed from an earlier, more widespread, and more primitive farming culture which perhaps evolved in the highlands of Guatemala. Be that as it may, the first true Maya communities, marked by characteristic art, architecture, calendrical system, and hieroglyphic writing, came into being, or at least first appear, in the low-lying Peten area at the base of the Yucatan Peninsula, where are found the ruined buildings of a number of the large religious centers commonly called cities. From an inferred Peten nucleus Maya culture seems to have spread more or less rapidly outward.

The era of development and early extension is known as the First or Old Empire; during that time the Maya reached their greatest intellectual and artistic heights. Eventually, however, all the southern Old Empire sites were abandoned; but the Maya, who, during the Old Empire, had been colonizing northern Yucatan, there carried on until their independent existence was terminated by the advent of the Spaniards in the first half of the sixteenth century. The later, northern, phase of Maya history is called the Second or New Empire. During the earlier part of this period there was continued cultural vigor, but toward the end civil wars, accompanied by military incursions from continental Mexico, brought marked retrogression, and it seems probable that even had the Conquest not occurred the Maya would never have re-attained cultural preeminence.

The above résumé of Maya history is based upon various sorts of evidence. Its chronological skeleton, however, is largely derived from the calendrical records of the Maya themselves. What we know of Old Empire chronology has been gathered from dates hieroglyphically inscribed upon the stone monuments or stelæ which were erected at many sites of that period. The dates in question are in the so-called Long Count, presently to be described. They cover the period between the Maya dates 8.14.10.13.15 and 10.3.0.0.0, a span of some 570 years. For New Empire chronology, on the other hand, we are principally dependent on the Books of Chilam Balam, post-Conquest documents written by the Maya in European characters, and containing dates which, although belonging to the same system as do the Long Count dates, are expressed in an abbreviated form, also shortly to be explained. To these abbreviated dates are referred a long series of events in pre-Columbian times, and some of the latest of the dates, concerning incidents connected with the coming of the Spaniards, are referable to years in the Christian calendar. Thus, were it possible to fix the Chilam Balam dates in the Long Count, we could, by working back from the just-mentioned terminal dates that are identifiable in terms of our calendar, determine the Christian equivalents of the Long Count records of the Old Empire. But this we have as yet been unable to do because of the abbreviated notation used in the Books of Chilam Balam. It is as if a Martian archæologist were to discover in Europe inscriptions covering the 570-year period 1042–1612; were also to come into possession of documents in which the mundane happenings of several centuries were referred to as occurring on such abbreviated or incomplete dates as '27, '56, '82, etc.; and were to know that the last of these dates, a year '92, signalized the arrival of his space ship upon the earth. His difficulty would be to identify the century to which that year '92 belonged; it might be 1792, 1892, or 1992. And so he would be unable to place, in his own calendar, the epoch between 1042 and 1613.

The efforts of students of the Maya to coordinate the Long Count and the Chilam Balam dates have reduced the currently favored correlations to two: that of Spinden, which may be called the 12.9.0.0.0 correlation, and which assigns to the Old Empire monumental record of 8.14.10.13.15–10.3.0.0.0 the Christian dates A.D. 68–629; and the Goodman-Martínez-Thompson, or 11.16.0.0.0, correlation, which brings the Old Empire dates forward 260 years to A.D. 328–889.

Either the 11.16.0.0.0 or the 12.9.0.0.0 correlation chronology may be correct to the very day, or either one may have missed the mark by over two and a half centuries. To clarify this seeming paradox and explain how the ambiguity arose in later Maya dates it is necessary to sketch briefly how the Maya calendar functioned.[2]

In our own calendar we have two concurrent reentering time cycles and a third non-reentering count, for when we write Sunday, January 1, 1939, we are recording positions in a 7-day cycle, a 365-day cycle, and also an infinite count of years. The Maya similarly had three counts, but their first, a reentering cycle, consisted of 260 days, while their third, an infinite count, was made in approximate years of 360 days. The second, also a reentering cycle and concurrent with the 260-day count, was of 365 days, but with a different monthly division.

The count of 260 days was formed of twenty names and the numbers 1–13, both repeating themselves in immutable sequence. Below is given a short section (65 days) to illustrate the system. It will be noted that after twenty days the same Imix and the rest of the sequence repeat themselves, but the cycle 1–13 has already completed one turn and now the number of Imix is 8 (1 + 20 − 13). Twenty days later the sequence is once more reentering with Imix, but now the attached number of Imix is 2 (8 + 20 − 26 [2 × 13]). At the next repetition we find 9 Imix.

TABLE 1. SECTION OF 260-DAY COUNT

(Read down left column and continue vertically with next to right.)

1 Imix	1 Ix	1 Manik	1 Ahau	1 Ben
2 Ik	2 Men	2 Lamat	2 Imix	2 Ix
3 Akbal	3 Cib	3 Muluc	3 Ik	3 Men
4 Kan	4 Caban	4 Oc	4 Akbal	4 Cib
5 Chicchan	5 Eznab	5 Chuen	5 Kan	5 Caban
6 Cimi	6 Cauac	6 Eb	6 Chicchan	6 Eznab
7 Manik	7 Ahau	7 Ben	7 Cimi	7 Cauac
8 Lamat	8 Imix	8 Ix	8 Manik	8 Ahau
9 Muluc	9 Ik	9 Men	9 Lamat	9 Imix
10 Oc	10 Akbal	10 Cib	10 Muluc	10 Ik
11 Chuen	11 Kan	11 Caban	11 Oc	11 Akbal
12 Eb	12 Chicchan	12 Eznab	12 Chuen	12 Kan
13 Ben	13 Cimi	13 Cauac	13 Eb	13 Chicchan

The number can no more be detached from the name than can a telephone number be detached from its exchange. Since 13 and 20 have no common factor, 260 days must elapse before the sequence starts to repeat itself. This calendar of 260 days was and, in some of the remoter Guatemalan highland villages, still is used for divinatory purposes. Shamans consult it for lucky and unlucky days governing every activity from planting crops to birth and marriage, the combinations of propitious or unpropitious numbers with equally fateful names supplying a complex but elastic divinatory apparatus.

The concurrent 365-day year consisted of eighteen months, each of twenty days, and five nameless days at the end. Since the only

---

[2] Detailed description of the Maya calendar is contained in Morley, 1915; for a simplified presentation, with bibliography containing the more recent contributions to this subject, see Morley, 1935.

common factor of 260 and 365 is 5, it is clear that only after the lapse of $\frac{260}{5} \times 365$ days, i.e. 52 vague[3] years, will the Maya equivalent of Sunday, January 1, repeat itself. A Maya date of that nature, such as, for instance, 1 Ahau 3 Zip (Zip is the third month of the series of eighteen, and 3 Zip the third day of the twenty) could, therefore, only occur again exactly at the close of 52 or multiples of 52 vague years. This period is called the Calendar Round; any date recorded in the two concurrent counts, such as 1 Ahau 3 Zip, is known as a Calendar Round date.

For the equivalent of our "1939 years after the birth of Christ," the Maya counted from a mythical date about 5000 years ago, employing for this the approximate year of 360 days and utilizing the vigesimal system (i.e. 1, 20, 400, 8000). Thus at the height of the Old Empire a date would be recorded as: 9 400-approximate-years, 16 20-approximate-years, 10 approximate years and, if there were any, so many odd days. As most dates recorded the ends of approximate years (called Tuns), we will assume 0 20-day periods and 0 days (note the vigesimal system also used here). The above, the Long Count, is written in our notation 9.16.10.0.0, and in this particular instance would have ended on the Calendar Round date 1 Ahau 3 Zip, referred to above. Such a system was as unequivocal as our Sunday, January 1, 1939, but it had the disadvantage of requiring for its presentation eight elaborate hieroglyphs, which were time and space consuming.

After 10.3.0.0.0 was reached no more dates were expressed in the Long Count, perhaps because of its clumsiness. Either the Calendar Round date only was recorded or else events were referred to the 20-approximate-year period (the Katun), this being denoted by the day on which it ended. The first system was less vague, but still not certain for reasons we shall now discuss.

The approximate year of 360 days and all its multiples in the vigesimal system, being divisible by 20, had to end on the same day name, which was actually Ahau. And as the 7200 (20 × 360) days which formed a Katun are divisible by 13 with remainder 11, it follows that the days, only 13 in number, on which a Katun could end ran in the following sequence: 1 Ahau, 12 Ahau, 10 Ahau, 8 Ahau, 6 Ahau, 4 Ahau, 2 Ahau, 13 Ahau, 11 Ahau, 9 Ahau, 7 Ahau,

---

[3] Note that *vague* year is used for the year of 365 days; *approximate* year for that of 360 days.

5 Ahau, 3 Ahau. Accordingly if a Katun ended on, for example, 13 Ahau, 260 approximate years later (13 Katuns) another Katun would end on 13 Ahau.

Such a Katun ended at 9.17.0.0.0, 10 approximate years after the Long Count date 9.16.10.0.0, 1 Ahau 3 Zip discussed above. The position of this and subsequent occurrences of Katuns ending on 13 Ahau are as follows:

A	9.17.0.0.0	(Katun 13 Ahau)
	13.0.0.0	(add 13 Katuns)
B	10.10.0.0.0	(17 Katuns + 13 = 30. Carry 20 to the next column in the vigesimal count, where it becomes 1 unit, leaving a remainder of 10.)
	13.0.0.0	(add 13 Katuns)
C	11. 3.0.0.0	(10 Katuns + 13 = 23. Carry 20 to next column, leaving remainder of 3.)
	13.0.0.0	(add 13 Katuns)
D	11.16.0.0.0	
	13.0.0.0	(add 13 Katuns)
E	12. 9.0.0.0	(16 Katuns + 13 = 29. Carry 20 to next column, leaving remainder of 9.)

We know from the Books of Chilam Balam that the foundation of Merida in January 1542, which marked the final conquest of Yucatan by the Spaniards, took place shortly after the end of a Katun 13 Ahau. The problem, therefore, is to decide which of the above Katuns 13 Ahau best meets the archæological and epigraphic evidence. As they are at intervals of roughly 260 years, a move from one position to the other similarly moves the Christian equivalents of Maya Old Empire dates by 260-year leaps.

A, the first position of 13 Ahau, can be eliminated without question, since definitely dated monuments were being erected in large numbers for sixty years after that Katun, and stylistic evidence shows that the dates on such monuments were contemporaneous with their erection, and cannot have been prophetic.

B and C can be eliminated on historical evidence if one assumes that there was no break in the calendar between the time when the Long Count was in force and the coming of the Spaniards. The rejection of these two possibilities, provided this assumption is valid, is due to the fact that, as has already been stated, the Books of Chilam Balam and other sources provide us with equivalents in

Christian chronology of Calendar Round dates of the time of the Spanish conquest. For example, the first day of the Maya year which started in the middle of 1541 was 13 Kan 1 Pop. Now a 13 Ahau, as we have seen, can end a Katun every 260 *approximate* years. The difference between 260 approximate and 260 vague years is 1300 (5 × 260) days, that is, a little over three and a half years.

Since 13 Kan 1 Pop is definitely correlated with the Christian calendar, each occurrence of a Katun ending 13 Ahau must be shifted the three and a half years, marking the difference between the two counts. The years in which the correlations automatically place the Katun 13 Ahau of the conquest are: B, 1546; C, 1543; D, 1539; E, 1536. As we know that the foundation of Merida took place in January 1542, and that that event occurred shortly after the end of a Katun on 13 Ahau, correlations B and C must be discarded or a break in the calendar assumed. The evidence against such a break is rather strong, but eventually archæological evidence against correlations D and E may be so overwhelming as to cause us to abandon the case against a break in the calendar. Once such a break is admitted, it is of course obvious that, so far as the epigraphic evidence is concerned, the Maya equivalent of the Spanish conquest may be anywhere after 10.3.0.0.0 and may not be connected with the end of a Katun 13 Ahau in the Long Count.

Solution of the correlation problem is naturally of vital concern to all students of New World prehistory. Most attempts toward that end have been based on the intrinsic evidence offered by the monumental inscriptions, the codices, and the post-Conquest records.[4] Success, along these lines, may eventually be attained. It is, on the other hand, conceivable that the inscriptional and documentary data can never be made to yield conclusive results. The present paper seeks to show that archæological methods give promise of making it possible confidently to choose between the several suggested correlations.

Archæological evidence, it is true, has already been utilized to a limited extent. At the very beginning of research upon Maya chronology, the fact that the southern cities were found in an advanced state of ruin, while those of Yucatan were far better preserved, some of them being still occupied at the Conquest, naturally led to a throwing back in time of the Long Count dates. Morley's brilliant reconstruction of Maya history and his concurrent attack upon the correlation question, while founded primarily upon the inscriptions, were buttressed by archæological observations. Spinden also has employed archæological reasoning in identifying as chronologically intermediate between the Old and New Empires certain ruins which lie north of the Peten cities and south of those in northern Yucatan. This he does in order to provide a home for the Maya during the long period which, under his correlation, must have elapsed after the abandonment of the Old Empire territory and before the relatively recent building of the great northern cities. He calls attention to changes in architectural practice, particularly in decoration, running northward from the Old Empire, through intermediate sites, and culminating in Yucatan. These he considers to mark the lapse of time. Other students have suggested that the observed differences may be the result of regional specialization.

---

[4] The most recent discussions of the correlation problem are: Roys, 1933; Palacios, 1933; Thompson, 1935.

The above are but a few examples of the many appeals to archæology which have been made by correlationists. But they have always been secondary, serving to support or controvert conclusions already reached in other ways. And it remained for Vaillant to open a new type of attack by reviewing the entire matter from the dispassionate viewpoint of the "dirt" archæologist. In a characteristically bold and acute contribution[5] he assembled the theretofore scattered data regarding the various stratigraphically determined ceramic sequences in the Maya and adjacent Middle American fields, combined them into a single general regional sequence, linked successive elements, or subperiods, of that sequence to known phases of Maya history, and, in a series of most illuminating charts, attempted "interarticulation of the ceramic sequences with the various correlations."

The results are of great interest. The pottery periods have to be stretched to inordinate lengths to fill the time required of them under the Spinden correlation; even for the shorter span of the Goodman-Martínez-Thompson they appear unduly long. Vaillant accordingly experimented with two new correlations: one (our C)

[5] Vaillant, 1935.

at 11.3.0.0.0, which brings the Old Empire Long Count dates up 260 years to 577–1148; the second (our B) at 10.10.0.0.0, advancing the Old Empire series still another 260 years to 837–1408.[6] The first (at 11.3.0.0.0) achieves much better accord than does the Goodman-Martínez-Thompson between ceramics and time; that at 10.10.0.0.0 yields a still closer correspondence.

These are revolutionary rearrangements. Even the more conservative 11.3.0.0.0 correlation requires that the Old Empire shall considerably have overlapped the New, while under that at 10.10.0.0.0 the Peten cities must have been in full flower until the very verge of the Conquest, and the Old and New Empires practically contemporaneous.

In concluding his paper, Vaillant emphasizes the fact that his charts are no more than trial balloons. He points out that further research will undoubtedly disclose new ceramic periods which, fitted into one part or another of the total sequence, will add to its length and perhaps bring it more nearly into conformity with the standard correlations. Certain such periods, as will be shown below, have indeed been identified since Vaillant wrote in 1935; there are strong indications of others.

Neither of the new correlations can perhaps stand the test of more comprehensive archæological data; their shortcomings from the epigraphic standpoint have already been pointed out in considering correlations B and C (page 502). But Vaillant's discussion is outstandingly valuable not only because it opens an entirely new avenue of approach, but also because it cannot fail to bring more closely together the excavators and the epigraphers, two groups whose members have hitherto not effectively pooled their efforts.

As has been intimated, we are now in possession of ceramic evidence not available to Vaillant in 1935. Studies by R. E. Smith of material recovered from Uaxactun by A. L. Smith show that the pottery of the Old Empire region in the Peten passed through four main phases, named by him Mamom, Chicanel, Tzakol, Tepeu; Mamom being the earliest, Tepeu the latest. J. E. Thompson's stratigraphic excavations at San José and other sites in British Honduras have confirmed the validity of this sequence and have supplied data indicating that the Tepeu phase can still further be subdivided and therefore probably lengthened.

---

[6] See also in this connection Thompson, 1935, p. 72.

These findings considerably expand the sequence which then seemed so hopelessly inadequate to fill the time required even by the Goodman-Martínez-Thompson correlation. They also serve to define the major phases of Peten pottery, and to give us accurate and detailed information regarding the many types of vessels produced in the Peten during the vitally important Old Empire period, information without which it would be impossible even to inaugurate the comparative studies and inter-area linkages that may eventually permit an archæological solution of the correlation problem.

The archæological solution here suggested depends upon our being able to accomplish two tasks: first, the dating of the pottery phases of the Peten in terms of the Long Count; and, second, the chronological equating of those phases with phases of the culture of southwestern United States which have been dated, by the Douglass tree-ring system, in terms of the Christian calendar.

The tying of Peten pottery to the Long Count would appear to be by far the easier undertaking. It presents, however, certain difficulties caused by uncertainty as to the extent to which the Old Empire Maya were given to shifting the locations of their stelæ. Had these monuments always been left where they were originally erected they would provide indisputable evidence for dating the pottery from such buildings, plaza fills, rubbish deposits, and tombs as can often be associated, stratigraphically or otherwise, with stelæ. But it is known that in some cases monuments were moved. If this reprehensible practice proves to have been common it will obviously be no simple matter to determine the exact Long Count position of the pottery phases. The complexities involved in this aspect of the research cannot here be considered. But it seems sure that further excavation carried on with this particular problem in mind can be counted upon to bring success. We shall therefore assume that the pottery phases can be dated in the Long Count and pass to the second step in the investigation, the chronological linking of the Peten with the ruins of Arizona and New Mexico.

Although, at first sight, it might be considered fantastic to attempt to bridge, with any degree of precision, the gap of over 1500 miles between Peten and the southern border of the United States, recent discoveries raise strong hope that it can indeed be accomplished.

Ceramic studies in all parts of the world have shown that the potter's art is never static. Styles arise, enjoy longer or shorter periods of popularity, and are succeeded by other styles. And ceramic, like biological, evolution does not lead to exact reproduction of earlier forms. Furthermore, pottery styles are limited, as far as manufacture is concerned, to definite areas. Parallel or convergent developments may take place but, again as in the animal kingdom, the results are apparently never such perfect duplications as to be indistinguishable. It is therefore possible, if we know our ceramic history well enough, to identify representative specimens both as to provenience and as to their age relative to other specimens. And because pots, particularly the more easily recognizable fine pieces, were commonly traded from group to group, sometimes over long distances, and because foreign types were often locally imitated, we are provided with much evidence for determining the chronological relationships between different cultures. Using such evidence, we may now take the first step in our long trek from the Peten to the Southwest.

At Kaminal-juyu, a large site in the outskirts of Guatemala City, the Carnegie Institution's excavations of 1936 and 1937 resulted in discovery of a series of tombs. The graves, evidently of persons of importance, were richly stocked with mortuary pottery. Among the vessels were a number of slab-legged, cylindrical-bodied tripods with lids, several of them stucco-coated and painted with elaborate designs in color. There were anthropomorphic vessels, the legs and lower body of the figure serving as base, the upper body, with arms and head, forming the cover. And there were also polychrome ring-base bowls with basal flange. All this pottery is evidently contemporaneous with the Tzakol wares of Peten: the slab-legged tripods and the effigies are very similar to Tzakol specimens from Uaxactun, and the polychrome bowls are seemingly importations from Peten.[7]

The tombs also contained several thin-walled, simple silhouette bowls of light orange-colored ware, decorated with straight and curvilinear incised lines and punctations; anthropomorphic vases with the features and arms represented by applied fillets of clay;

---

[7] When the technological methods under development by Miss Anna O. Shepard have been applied by her to the Kaminal-juyu material, we may learn the actual source of some or all of the above specimens. But whether or not they prove to have come from Peten, their resemblance to Peten forms indicates a close relation in time. In this study we are not concerned with the direction of artistic influences, merely with problems of chronology.

and a polished black tripod, the lines of its elaborately carved and incised decoration filled, after firing, with brilliant red paint. These pieces are practically duplicable in the collections made by the Swedish archæologist Linné at San Juan Teotihuacan near Mexico City.

Contemporaneity of the Teotihuacan culture and that of the Maya Old Empire has for some time been suspected,[8] but not until discovery of the mortuary vessels from Kaminal-juyu could it be considered proved. And this pottery permits the further step of correlating with Teotihuacan a definite phase of Old Empire ceramics, the Tzakol.

The Kaminal-juyu finds have most significant implications. They make it clear that the people of the Guatemala highlands were in close touch with groups in Central Mexico. In fact, Mexican rather than Maya influence seems to have been preponderant in that region, at least during the period of the tombs that were excavated. But they are of more immediate importance for the present study.

Under the Spinden correlation, which ends the Old Empire at about A.D. 628, it was necessary to believe that the Teotihuacan culture, roughly dated in the legendary Annals of Quauhtitlan as having existed from A.D. 750 to 1050, was a post-Old Empire and probably an Old-Empire-stimulated development, whereas we are now certain that Teotihuacan and Old Empire were more or less closely parallel growths. So, if the Annals of Quauhtitlan are even approximately correct the Spinden correlation must be discarded.

The new data from Guatemala cannot, however, be of maximum value, either for throwing light upon Middle American culture relationships or for solving the correlation problem, until we can determine to which phase of the Teotihuacan culture the Kaminal-juyu specimens should be referred. This the writers are unable to do, their first hand knowledge of Mexican ceramics being very slight, and the several Teotihuacan phases into which Teotihuacan pottery is apparently divisible having not as yet been adequately described. But probably on the basis of material already in the hands of Mexican archæologists, and certainly as a result of further stratigraphic work in Teotihuacan sites, a clean-cut hook-up can be made. When that has been accomplished we shall be ready for the next northward advance. And while one can hardly look forward to leaps so rapid or so long as the one which took us, via Kaminal-juyu, from the Peten to Teotihuacan, and although the evidence may be at first sight less clear, we are confident that it will eventually be possible to link the Teotihuacan ceramic phases to those of Arizona and New Mexico, where, it will be remembered, the Anasazi cultures have been given absolute dates by Douglass and his pupils. This most desirable result depends upon investigations in the area between the Valley of Mexico and the southern border of the United States.

---

[8] For the latest discussion of this point see Krickeberg, 1937, p. 212.

The archæology of northern Mexico is as yet poorly understood. Ruins, however, are numerous and studies now in progress, notably those of Sr. Noguera of the Mexican Dirección de Monumentos Prehispanicos, Drs. Sauer and Isabel Kelly of the University of California, Dr. Brand of the University of New Mexico, and Mr. Eckholm of the American Museum of Natural History, can be counted upon to establish the necessary chain of relative cross-datings. It is also quite possible that absolute dates may be supplied, by dendrochronology, for some of the more northerly Mexican cultures, thus shortening the gap that must be bridged by archæological methods.

By such means, we believe, data are obtainable which will allow us to decide which of the several correlations between the Maya and Christian calendars is correct. Final conclusions must obviously not be based on single objects or even upon the first apparently successful tying of Teotihuacan to Anasazi. It might still be objected that cross-dating by ceramic linkages is not precise enough to establish a point so vital; for cultural influences, as reflected by pottery, may travel at varying rates; a number of cultures will probably have to form part of the chain, and cultural lags in peripheral regions can be misinterpreted; specimens obtained in trade, and retained as heirlooms, might lead to false deductions regarding the time relations of sites. If the correlations differed from each other by a shorter time, as could be the case if there was a break in the Maya time-count (see page 503), the suggested method might well be mistrusted. It would not allow us safely to choose between correlations fifty or even a hundred years apart. But the probable span being two full centuries and a half, we have so great a leeway that the admitted chronological imprecision of archæological cross-dating can, we are certain, be discounted.

LITERATURE CITED

DOUGLASS, A. E., 1935. Dating Pueblo Bonito and Other Prehistoric Ruins of the Southwest. *Nat. Geog. Soc., Contributed Technical Papers, Pueblo Bonito Series* No 1.

KRICKEBERG, W., 1937. Bericht über neuere Forschungen zur Geschichte der alten Kulturen Mittelamerikas, *Die Welt als Geschichte*, vol. 3, nos. 2–3, pp. 194–230.

MORLEY, S. G., 1915. An Introduction to the Study of the Maya Hieroglyphics. U. S. Bur. Amer. Ethnol. Bull. 57, Smithsonian Inst.

——— 1935. Guide Book to the Ruins of Quirigua. *Carnegie Inst. Wash. Supp. Pub.* No. 16.

PALACIOS, E. J., 1933. El Calendario y los Jeroglíficos Chronográficos Mayas. Mexico, D. F.

ROYS, L., 1933. The Maya Correlation Problem Today, *Amer. Anthropologist*, n. s., vol. 35, pp. 403–417.

THOMPSON, J. E., L. ROYS, and R. C. E. LONG, 1935. Maya Chronology: The Correlation Question. *Contr. Amer. Archæol.*, no. 14, Carnegie Inst. Wash. Pub. No. 456.

VAILLANT, G. C., 1935. Chronology and Stratigraphy in the Maya Area, *Maya Research*, vol. 2, no. 2, pp. 119–143.

RACE AND CLASS IN YUCATAN

ROBERT REDFIELD

University of Chicago
Research Associate, The Carnegie Institution of Washington

The language of every society has terms for denoting the races and classes that matter to its members, and in the development of these terms may be read something of the history of that society. Where two ethnic groups occupy the same or adjacent territories, or where a society is made up of classes or castes, there are local words to distinguish them, and in the connotations of the words are to be found the attitudes by which convention has defined the position of the groups with reference to one another. "Barbarian," "helot," "villein," "lord," "Yankee," "greaser," "goy," and "Jew" might serve as gateways to an immense amount of history of social relations. And what significant changes in the societal position of people of African origin are briefly recorded in three forms of the same word: negro, nigger, Negro!

From this approach one may consider the history and the outcome of Spanish-Indian contacts in the peninsula of Yucatan. This paper deals with certain aspects of these contacts as they appear in the meanings attached to the four terms *mestizo*, *vecino*, *mazehual*, and *dzul*. It is the meanings attached to the terms at the present day, or in times within the memory of natives of Yucatan still living, that necessarily provide most of the materials because the author and his associates,[1] as ethnologists and sociologists, have dealt with the present-day society of the peninsula. But questions raised by study of present conditions call for answers out of knowledge of past conditions. Here the writer will turn, when he can, to such glimpses as have been afforded him by occasional contacts with students of colonial and ancient Yucatan.[2] They are, however, only glimpses, and most of the questions will remain unanswered until the sources of information on the early periods are more thoroughly studied with reference to the problems mentioned below.

---

[1] Dr. Asael T. Hansen and Mr. Alfonso Villa. These gentlemen have made large contributions to my understanding of the matters discussed in this paper. The statements regarding the "old society" in Dzitas are based largely on abundant materials secured from informants by Margaret Park Redfield.

Four communities have provided our materials: a group of villages of tribally organized Maya Indians dwelling in the remote isolation of east central Quintana Roo; a village (Chan Kom) of peasant Indians a few miles south of the well-known archæological site of Chichen Itzá; a town (Dzitas) of mixed-bloods and Indians partly agricultural, partly commercial, situated on the railroad just north of Chichen Itzá; and the one city and provincial capital, Merida. The reasons for this selection and the general characteristics of the four communities have been set forth in an earlier publication.[3] The second community, the peasant village, has been described at length.[4] For material on the other three communities I have looked to unpublished notes and manuscripts: those of Mr. Villa in the case of the Quintana Roo villages, those of Dr. Hansen and Mrs. Hansen in the case of Merida, and my own and Mrs. Redfield's in the case of Dzitas. Many oral exchanges of fact and opinion among this group of workers have contributed to the production of this publication, as of the others.

I

If the few Negroes, Chinese, Koreans, and Europeans are disregarded, there are in Merida just two contrasting kinds of people which the local community clearly defines: *mestizos* and *gente de vestido.* The terms "rich" and "poor," "upper class" and "lower class" are not precisely denotative; there are many people who are neither rich nor poor, and many whose class position is uncertain. As we shall see, the terms "Indian" and "White" do not denote significant groups in Merida. But with *mestizo* and its opposite, *gente de vestido*, there is practically no doubt because the words have reference to contrasting kinds of costume, and every individual either is wearing one of the two costumes or is wearing the other. If a man wears cotton trousers, cotton shirt, and especially sandals, he is a *mestizo*; if a woman wears the *huipil* and *rebozo*, she is a *mestiza*; and it does not matter what be the color of their skin. Correspondingly any wearer of modern European clothes is *de vestido*.

---

[2] I have especially made use of information and suggestions generously provided by Mr. Ralph L. Roys.

[3] Robert Redfield, Culture Changes in Yucatan, *Amer. Anthropologist*, vol. 36, no. 1, pp. 57–69, at pp. 62–64, 1934.

[4] Robert Redfield and Alfonso Villa, "Chan Kom, A Maya Village." Carnegie Inst. Wash. Pub. No. 448, 1934.

This distinction is recognized almost everywhere in Yucatan. The basic terms by which Yucatecan society classifies its members refer not to race, nor to genealogy, nor to wealth, but to costume. The costume is, however, connotative of class position. *Mestizos* are members of the lower class; costume is the paramount symbol of lower class position. If you see a *mestizo* costume you see someone who may be a laborer, or a house servant; he could not be a lawyer, or a business man, or even a street car conductor. The basic terms *mestizo* and *de vestido* recognize an essential division of society into two classes: the élite and the non-élite, the classes and the masses. These terms imply a stable social order so constituted. The children of *mestizos* should be *mestizos*; the children of *gente de vestido* are expected to have children who wear European dress.

But the society is not stable. The *mestizo–de vestido* distinction is not as important as it is remembered to have been. The society is undergoing great change. In Merida especially the *mestizos* are giving up the folk attire and putting on European dress. They are doing it so rapidly that the folk costume is likely to become extinct, at least in the city, in a few decades. The degree to which the folk costume has been given up is evidenced by the fact that today while there are only four social clubs for *mestizos*, there are at least twenty-eight for the recently *de vestido*. Had we no other knowledge of the change we should nevertheless be informed of it from the fact that there are derogatory terms (*cuch-vestido*, *dzul del pueblo*, *dzul falso*) for such costume changers. To conservative *mestizos*, proud of their class and the garb which is its badge, such persons are to be condemned. "To the conservative," writes Dr. Hansen of Merida *mestizos*, "the *vestido* garb has a semi-sacred quality. It is entirely inappropriate in the type of life the lower class must lead. To carry a basket in the market or to wash clothes while wearing the *vestido*, almost defiles the costume."

Nevertheless the change goes on. European garb is the symbol nowadays not so much of the status of a privileged class as of opportunity in a freer and more individuated society. It is identified with education and a better job. If a *mestizo* lacks the courage to make the change in costume himself, he can make it through his children by putting them in European dress. If the question be asked directly, people will often reply that there are *three* classes in Merida; they will describe these three classes in terms of occupation and wealth, and they will recognize a middle class composed of people fallen from the old upper class and of new arrivals from the old lower, *mestizo* class.

Today there are many *de vestido* people in humble occupations. To say "we *mestizos*" means less than it did; the tendency nowadays is to say "we poor people." The old-established upper class person, on his part, is less likely to use *gente de vestido* to denote his class; so many pushing people claim the term now and with it identify themselves with the old aristocracy. So the terms with which Merida in fact describes itself today are many and vague: the upper class, the middle class, the lower class; the poor and the rich; the professional and business and other occupational groups; the *mestizos*, the *gente de vestido*, and the people-who-used-to-be-*mestizo*-but-are-now-*de-vestido*-or-are-on-their-way.

The terminology now in use makes known to us a society that is in process of change, in which social position is in considerable degree a matter of individual accident, choice, and initiative. But the basic pair of terms, impaired though they are in definition, record a condition of society in which two status groups, almost like castes in their fixity, composed the society of Yucatan.

This two-class society existed in its full vigor, apparently, just before the memory of the oldest natives now living. Dr. Asael T. Hansen and Sr. J. Ignacio Rubio Mañé have investigated some of the documentary sources on Yucatan in the early decades of the nineteenth century. With Dr. Hansen's permission I quote from his unpublished manuscript on the subject. Intra-class movements, he says, "rarely led to crossing the boundary between the two main classes. No matter what reverses a person in the upper grouping suffered, he remained in that category. Not a single example to the contrary was ever encountered. The line was not absolutely inviolable from the other direction." In six decades of the nineteenth century not more than a few dozen *mestizos*, having amassed wealth, abandoned the folk costume. "All of these cases were isolated achievements. They probably did not blur the line in the least." *Mestizos* "accepted their general place in the social scale without question. They tended to rank themselves with reference to other members of the category rather than with reference to the total range of status. The goal that the ambitious sought was to attain an enviable position in the lower class; to be able to do lower class things that were out of the reach of most of their fellows. Such an accomplishment brought complete satisfaction. Upper class people could strive for other ends; ends that members of the lower class recognized as higher than their own." In such a society, where the status of every individual was sure and was obviously announced in his costume, "superiors and inferiors could rub shoulders without danger of offending either, and they did so in patron fiestas and on many other occasions." Norman, writing in 1841,[5] was struck with the easy intimacy of the social classes. It is apparent that the connotation of *mestizo* as used by a conservative older person of Yucatan reaches back to a state of society which the historical investigations of Sr. Rubio Mañé and Dr. Hansen recall for us.

We are not yet done with the word *mestizo.* There is a puzzle written on its face. Why should a word which ordinarily refers to a person of mixed racial ancestry, in present-day Yucatan mean in nineteen uses out of twenty a person wearing a certain sort of costume? This question reminds us of others: What has become of the Indians? Who, in Merida today, is "an Indian"?

An answer to the last question leads toward an answer of the first. To the Meridano the Indians are, generally speaking, not to be found in Merida, but in the villages and settlements out in the bush. The word *indio* is, however, sometimes applied to those relatively few city dwellers who wear the folk costume and who bear Maya surnames. Such persons are included among the *mestizos* in the common general meaning of the word. But in a less common and special sense a *mestizo* is a person who wears the folk costume and who bears a Spanish surname. Although the difference between a person with a Maya surname and one with a Spanish patronymic is of little importance in Merida today, yet one who wishes to speak depreciatively of another may call him an Indian, and if the latter bears an Indian surname he cannot cast off the epithet. The distinction, on the basis of surname, between two stratified divisions of the lower class is a mere vestige in Merida.

---

[5] B. M. Norman, "Rambles in Yucatan," p. 28. New York, 1888.

Once the distinction must have been clear. The early lists of population[6] recognize three groups: Indians, *mestizos*, and Whites (Spaniards). The very existence of the word *mestizo* testifies to the previous recognition of a threefold division of the population along ethnic lines. When Indians mixed with Whites it became necessary to call the marginal individual by some term. *Mestizo*, in the primary sense of racial hybrid, was applied. The problem arose, inevitably, in any situation of racial intermixture, of the social position of the mixed person. Should he be identified with the Indian group? Should he be identified with the Spanish group? Should he be recognized as constituting an intermediate social category? Without the benefit of particular knowledge of historical fact it is nevertheless safe to say that all three of these results took place. The offspring of casual unions between Spanish man and Indian woman, unrecognized by the father and hence bearing an Indian surname, were brought up with Indians and must have been identified with the Indian group. If certain such individuals were marked out as having Spanish blood, the tendency to do so must have declined with the spread of white admixture throughout the socially Indian group. On the other hand the offspring of legitimate marriage between Spaniard and Indian (of which there were a few) must, at least in some cases where the mother was an Indian of prestige and where the Spanish father gave the children education, have been identified with the Spanish group. At the same time many offspring of nonlegal but enduring unions of Spaniard and Indian involving the setting up of a domestic establishment must have occupied an intermediate position. The father, associated with his offspring in familial intimacy, could not repudiate them entirely; frequently, we know, he allowed them to take his surname. Yet as illegitimate halfbreeds, raised by Indian mothers, they could not, in most cases, be identified with the Conquerors. Thus there came about a socially intermediate group of racial hybrids. Some children of formal intermarriages and some resulting from casual union also probably came to be associated with the intermediate group.

---

[6] A census of Merida in 1790 reported 14,751 Indians in Merida. Estado General de la Población de la Jurisdicción de Mérida, Capital de la Provincia de Yucatán. Año de 1790. General y Público de la Nación, Mexico, D. F., Historia, tomo 522, exp. 257. Published in *Discurso de las Provincias de Yucatán y Campeche, Documentos para la Historia de Yucatán.* Merida, Yucatan, 1938.

At this period—during, let us say, early Colonial times—a *mestizo* was, we may suppose, a person of mixed Indian-Spanish ancestry who wore folk costume but who bore a Spanish surname. In the former character they were like the Indians; in the latter like the Whites. Together Indians and *mestizos* constituted the lower classes, but they were discriminated, Dr. Hansen tells me, in both the civil and the ecclesiastical law: there were certain differences as to religious privilege; *mestizos* were under the direct municipal government of Merida, while Indians had their own municipal government subordinated to Spanish authority; *mestizos* but not Indians were obliged to perform military service, while *mestizos* were free from obligatory labor of all kinds; the Church discouraged intermarriage between the two groups. There were also certain distinctions recognized by custom alone, as in details of the folk costume and in the division of labor between the two groups. But in important essentials Indians and *mestizos* were alike: they were the common people, doing the common labor and accepting lower status; and they wore essentially the same costume.

The word *mestizo* must have come to mean primarily a member of the lower class wearing the folk costume as the line between Indian and *mestizo* became blurred. Dr. Hansen puts it thus: "The development may be viewed as a linguistic response to the fading of the Indian-*mestizo* demarcation and as further evidence that the basic status groups were White and non-White or what we have called the upper class and the lower class." The number of *mestizos* increased relative to the number of Indians. In spite of restrictions against marriage between Indians and mixed-bloods with Spanish names, many such marriages occurred. Sr. Rubio Mañé tells us that in the eighteenth and nineteenth centuries a practice flourished whereby upper class fathers had their illegitimate children by Indian mothers baptized as foundlings. "In some years," says Dr. Hansen reporting on Sr. Rubio Mañé's investigations, "foundlings made up more than one-fourth of the total baptisms in the cathedral of Mérida—the church of the whites." In certain cases white godparents gave their surnames to

Indian godchildren. In these ways the *mestizo* population multiplied. At the same time differences in costume between Indian and *mestizo*—they had never been great—gradually disappeared. Dr. Hansen's references to conditions in the early part of the nineteenth century indicate that by that time the two groups had the same occupations and the same dress. The legal restrictions on the Indians were revoked or fell into disuse after independence from Spain. By the middle of the nineteenth century the mixed-bloods with Spanish names and the Indians—now also largely mixed in blood—with Indian surnames looked alike, dressed alike, and were treated substantially alike. There were many more of the former than there were of the latter, and it was easy to apply the term earlier reserved for this subgroup to the entire, uniform-wearing lower class. A word for a biological hybrid had come to designate the lower of two social classes. This coalescence of mixed-blood and Indian groups suggests the manner in which mulatto and Negro came in the United States to form a single caste-like class, but with the important differences that Yucatan, like the rest of Latin America, drew no color line, and that the badge of the Yucatecan class was a costume, not a skin color. And now it comes about that with the breakdown of other features of the old social organization of the penninsula, and with the freedom of the individual to give up the folk costume and no matter what his birth to compete with others for status, the class lines tend to disappear.

II

We have seen that in Merida there is now a shifting three-class system with much individual mobility. At the same time the terms *mestizo* and *de vestido* represent a two-class system that no longer completely exists. Some reasonable speculations, developed from consideration of the word *mestizo*, supported by a little historical data, led us back to a condition of Yucatecan society in which there were three ethnically defined classes, or, perhaps it would be better to say, in which the lower class was divided into two stratified parts: the people with Indian surnames and the people with Spanish surnames. Of this class division there is in Merida today hardly a trace. In Yucatan, however, if one wants to find an older condition of society represented currently, all he needs to do in many cases is to go farther away from the one city into the rural communities. One can go back either in time or in space and, speaking broadly, arrive at the same condition of society.

The people of the little town of Dzitas, like those of Merida, exhibit every shade and combination of Indian and Spanish racial intermixture. There is relatively more Indian blood in the population, and much more Maya is spoken. There is the same difference between *mestizos* and *gente de vestido* that is recognized in Merida, but there are proportionately more *mestizos* than there are in the city. In general the people who dwell on the outskirts of the town are Maya-speaking, *mestizo* people who carry on a simple agriculture. In news from Dzitas sent to the Merida newspaper by the local correspondent these people are referred to as "members of our working class." The writer of such notices, like others of the prestige-enjoying people who have their homes near the central plaza, is a Spanish-speaking *de vestido* person with a specialized occupation. Included in his group are such specialists as carpenters, masons, storekeepers, and barbers. These occupations are open to anyone, and many persons apparently of all-Indian ancestry are to be found in them. Dzitas, it soon appears, recognizes two social classes, the lower and middle class of present-day Merida; the outlines of the classes are ill defined and ethnic origin or pedigree does not determine class membership, a combination of factors determining the social position of the individual.

This is, however, only the apparent, the public, class structure. Behind it exists another, remembered rather than actual. People in Dzitas, especially older people, divide their neighbors into two groups: *vecinos* and *indios*. A *vecino* is anyone with a Spanish surname. An *indio* is anyone with a Maya surname. The older distinction to which the conservative people still cling is that between the Peches, Poots, and Nahs on the one hand and the Leals, Arceos, and Aguilars on the other. The former are, or should be, the lower class; the latter, the upper class. Any *indio* is theoretically inferior to any *vecino*. No matter how poor or how dark-skinned a man or a woman may be, or how humble his occupation, if he has a Spanish surname it is proof that he is not an Indian, that he is a descendant of Spaniards, and that *indios* should recognize that he is better than they. One case, well known to the writer, illustrates the force and tenacity of this old genealogical class divi sion. The case is that of an old man, dark-skinned as any Indian, a wearer of the folk costume. He speaks no Spanish; he carries on primitive agriculture like any of his poor neighbors, and he participates in the pagan agricultural ceremonies. But he has a Spanish surname; therefore he looks down upon his neighbors with Maya names. On one occasion he came to the plaza and watched the activities of the municipal government—since the Revolution in the hands chiefly of Indians—"to see," he said, "what a bad job those Indians are making of the town government."

As may be supposed, the *vecino-indio* distinction is emphasized especially by *vecinos*, and still more specially by *vecinos* who have lost the other indicia of status: those who are poor, who have no longer specialized occupations, and who cannot afford *vestido* attire. A person with three *indio* grandparents and one, the father's father, a *vecino* grandparent clings to the prestige attached to the Spanish surname. A woman, impoverished to beggary, with four *vecino* grandparents does her best to look down upon her well-to-do *indio* neighbors, is scornful of *indios* who get into trade or take houses near the plaza, and complains of the fact that *indios* do not show her respect. Although marriages between *indios* and *vecinos* are fairly common, and informal unions are even commoner, nevertheless when an *india* becomes the wife of a member of a *vecino* family that maintains some cohesiveness and class pride, she is not readily accepted, and the bride's *indio* pedigree may be a matter of embarrassment.

It is clear that the *vecino-indio* distinction is that between the two levels of the fundamental Yucatecan lower class. It is the distinction which we have seen was once more clearly recognized in Merida, but which is there practically extinct. In the thinking of the older generation in Dzitas it is *the* important social distinction. Either the upper white class was never clearly represented in Dzitas, or it early disappeared. The *vecinos* are the upper division of the old lower class; but in old Dzitas they were the highest class that local society recognized. A division was formed between the Indians, on the one hand, and the mixed-bloods on the other. With intermixture, surname came to be the essential symbol of class membership.

Apparently the old class structure still flourished in the youth of the oldest natives of Dzitas now living. They give many details of the conditions which then prevailed. Only *vecinos* lived near the plaza, and all the *indios* lived on the outskirts of the town. The posts in the local government were filled, with only the rarest exceptions, by *vecinos*. The *indios* had their own separate municipal government, subordinated to that of the *vecinos*. The *indios* met in a separate place of public meeting, called *audiencia*. The decisions and orders of the *vecino* government were transmitted to them through their own chieftain (*batab* or *alcalde*) whom they themselves selected. Only a *vecino* might be a soldier; in military activities the *indios* were permitted to be watchmen and messengers. In short, the *vecinos* were politically dominant; they were the real citizens of Dzitas; the *indios* maintained a separate government for matters concerning exclusively *indios*; and the *vecinos* ruled the *indios*, making use of the *indio* officers.

The *indios* occupied positions of less dignity and acknowledged their inferior position in name, costume, occupation, and place of residence. An older *indio* was addressed with a term of respect, recognizing his years, but the term (*yum*) was not the same by which a *vecino* was addressed; only *vecinos* were called *don*.[7] By the manner of salutation the *indio* expressed deference, while the *vecino* received it. In domestic life both classes wore the folk costume, but there were many small differences in costume associated with the difference in status. The *huipils* of the *indias* had little or very simple embroidery, while the *huipils* of the *vecinas* (*ternos*) had cross-stitch embroidery. "If an *india* had worn a *terno* it would have been taken from her and burned at the door of the church." The *vecinas* wore petticoats, coral hairpins, fine combs, and head scarves of fine cloth; the *indias* had to content themselves with hairpins of native tortoise shell, wooden combs, and head scarves of crude cotton cloth. It was the *vecinas* who responded to the fashions. They went to mass in the costume of the city (*de vestido*), and as each novelty in dress or display reached Dzitas—gold chains, good shoes, new embroidery designs—it was the *vecinas* who took it up, and public opinion allowed the *indias* to assume these modes only when they had been abandoned by the class above theirs. It is said that in those times the Church discouraged intermarriage between the classes by charging a higher fee for such marriages, and indeed the books of the Dzitas civil registry, which go back to 1867, tend to confirm this assertion.

---

[7] Nothing was told the writer in Dzitas of Indians who had been made *hidalgos*, as we know was done not infrequently in Colonial times.

An examination of the distribution of cultural, linguistic, and occupational characters among the *vecinos* and *indios* of present-day Dzitas reveals the discrepancy that has developed between the old ethnic classes and the present more flexible status groups. Out of 225 households of Dzitas, thirty-nine are of Spanish-speaking, *de vestido* people carrying on specialized occupations. These people are with very few exceptions *vecinos.* At the other extreme are 136 households of Maya-speaking, *mestizo milperos* and woodcutters. All these people, in the thinking of the conservative native, ought to be *indios*; in fact, most of them are; the minority that are *vecinos* tend to regard their position as an anomaly. Furthermore (besides 12 households for which data are lacking), there are 38 households in which the social factors are inconsistently distributed; included are Maya-speaking *mestizos* who have specialized occupations, Maya-speaking *de vestido* people, and Spanish-speaking people, with the woman only *de vestido*, some of whom are simple agriculturalists and others of whom carry on specialized occupations. Of these intermediate families some have Spanish names and some Indian names. In the thinking of the old conservative these "mixed people" should not be there at all. The Indians have "got out of their place." They control the municipal government; they are becoming barbers and brakemen; they put on city clothes; they join the social club; they do not, in gesture or speech, show respect for *vecinos.* Meanwhile some *vecinos* have lost power, grown poor, and given up their Sunday *de vestido* attire. Left behind in the changing world, bereft of the symbols of the status which traditionally they should have, save the Spanish name, they emphasize the worth of this and cling to the memory of a definite class society in which membership was decided at birth.

As in the case of the word *mestizo*, there is a question as to how *vecino* ("neighbor," "citizen") came to denote a member of a social class. In Merida today the word means "neighbor"; it is not used there to mean a non-Indian, although people in Merida know that in such places as Dzitas it does have this meaning. Formerly, however, according to Dr. Hansen, "at least from 1847 to 1881, the population of Yucatan was legally divided into two categories, Indians and *vecinos*." It may be that the term *vecino*, though recognized in law, was never popularly employed in Merida for the higher division of the lower class because in Merida there was an important upper class of white aristocracy who of course also enjoyed all civil rights. *Vecino* meant non-Indian and so included two groups socially distinguished: Whites and mixed-bloods. But in such a place as Dzitas, where this supreme class was unrepresented, the mixed-bloods with Spanish names were the only ones who enjoyed full civil, military, and ecclesiastical rights; they were complete citizens, in a sense in which the Indians were not. So the word *vecino* came, especially in such communities, to denote the upper of the two classes.[8]

The situation in Dzitas may be summarized. The society maintains in a moribund condition a two-class structure. These classes correspond to the division within the lower (*mestizo*) class in Merida, where the division is practically lost. The Dzitas classes are defined as Indians and non-Indians, but as assimilation and amalgamation had distributed racial features of physiognomy throughout the population and developed a single culture for all, the surname, a certificate of ancestry, had become the symbol of class position. Nevertheless these ethnic or genealogical classes, no longer supported by legal provisions and undermined as a result of freedom of economic and educational opportunity for all, exist more in retrospect than in current fact. Their place is taken by a vague distinction between "the working people" and "the best people" (there are no accepted terms for these groups). The status of the individual is not determined by his birth; he may seek it and make it for himself.

---

[8] Mr. Ralph L. Roys tells me that, in early colonial documents that he has seen, Indians are rarely referred to as *vecinos*. Yet he has noted such instances in one document of the seventeenth century and another of the eighteenth, and so we must conclude that at least as late as 1788 (a petition in the Xiu papers, dated December 13, 1788, Massachusetts Historical Society Reproduction, p. 148) the word *vecino* could be applied to persons with Indian surnames. On the other hand Dr. Hansen calls my attention to a publication of the government of Yucatan in 1848 which gives the requirements for candidacy for justice of the peace. The statute provides that to hold such a post one must be a literate, married citizen over twenty-five in full possession of civil rights; but if such persons are not available, then unmarried persons may be chosen; but "a falta absoluta de vecinos, podrán ser nombrados los indígenas que siempre reunan las expresadas circumstancias, y en este caso serán desde luego considerados en la clase de los antiguos hidalgos." There are not a few eighteenth century documents in which non-Indians are distinguished as *vecinos*.

III

In Merida even the older class system (*mestizos* vs. *gente de vestido*), while it has some ethnic implications, is defined in terms of costume. The older class system of Dzitas is in terms of the distinction between Indians and non-Indians. The difference between Indian and non-Indian, or between White and non-White, became with the Conquest the most important group difference. In Merida the distinction is almost gone. In Dzitas it remains in the ethnic or genealogical significance attached to two decaying classes. If we wish to find the White-Indian distinction more completely represented, we must leave the railroad towns with their mixed populations and their direct contacts with the capital, and go to villages south and east of Dzitas. One such, Chan Kom, is an all-agricultural settlement of about 250 people situated a day's walk from Dzitas. This village is like many another in the region in that its population, Maya in speech and homogeneous in folk customs, is composed of tax-paying, school-attending peasants who regard themselves as part of the state of Yucatan and participate in its formal government. Beyond a belt of uninhabited territory, three days' travel to the southeast in east central Quintana Roo, lie other villages of similar blood and culture, but composed of tribally independent natives who have never, since the War of the Castes in the middle of the nineteenth century, made complete peace with the government of Mexico.

Inside the village—any of these villages—there are no social classes. Status is a matter of sex, age, chosen occupation, and personal qualities; it is not ranked. So there are no words for class differences inside the local communities. The important distinction for these villagers is that between the people who belong to the village or the group of neighboring villages maintaining strong solidarity on the one hand, and outsiders, people who are not "of us," on the other. And for this distinction two Indian words are chiefly used: *mazehualob* and *dzulob*. Generally speaking, a *mazehual* is one of us, an Indian in blood and customs, while a *dzul* is an outsider, a white man with different customs. But while these general statements may be validly made of Chan Kom and of Tusik in Quintana Roo alike, the situation in its details presents differences that require a separate treatment.

In Chan Kom the word *dzul* implies respect. It is applied to any person who wears European clothes, who exhibits in conspicuous degree a light skin and Caucasoid features, city ways and education, and whose surname is not Maya. No one who wears the folk costume is "a genuine *dzul*," and no one who has a Maya surname. But if a man lives the life of the city and is educated and authoritative, he may be called *dzul*, even if his skin is dark. For persons who are not complete *mazehualob*, but who nevertheless fail to exhibit the full characters of a *dzul*, the phrase *kaz-dzul*[9] is used. It ordinarily denotes an outsider to the local community. A visiting trader or other townsman, who is *de vestido* but who leads a rustic life or exhibits Indian physiognomy, may be called a *kaz-dzul*. The word may be applied to someone who leads the life of the village but bears a Spanish surname. The term *kaz-dzul* is mildly depreciative. A man will not easily apply it to his fellow villagers (there are only a very few bearers of Spanish surnames in Chan Kom); if his attention is called to the fact that his fellow villager has a Spanish surname he will admit that then the latter cannot be quite fully a *mazehual*. But he ought to be.

The word *mestizo* may be used in Chan Kom, but uncommonly, for any *kaz-dzul* who has a Spanish surname and who wears the folk costume. It therefore does not have the connotation of a social class which it has for the Meridano; it implies a kind of folk costume shared by the speaker and a kind of surname which the speaker, a *mazehual* or *indio*, does not share, but it also carries with it the notion of a mixed or intermediate person. The word *vecino* is rarely used, but it is known as an equivalent to *mestizo*. Thus the basic meanings of both words are the same as those which attend them when they are spoken in Merida or in Dzitas; yet in the contexts of the local societies of the villages they play a much smaller and a different part.[10]

While the people of Chan Kom do recognize *kaz-dzulob* among their number, yet it is apparent that the solidarity of such a village as Chan Kom is such as to make difficult an admission that the villagers are not all one single ethnically homogeneous people, "we, the *mazehualob*." Even though the native looks up to the *dzul*, he thinks of the *dzul* as outside his local group, and it is not easy for him to think of even a *kaz-dzul* as a full member of the local society. Some incidents illustrate the point. A *kaz-dzul* trader from a town wishes to settle and open a store; he does not receive the welcome that a *mazehual* from a friendly neighboring village would receive. A native of the village, affected by city ways and seeking prestige in the townsman's terms, proposes to give up his Maya surname for a Spanish surname; the villagers disapprove. This tendency to identify the native participant in the village life fully with the in-group is even more marked in Quintana Roo. There it is known that three generations ago white men, renegades and captive, settled in the villages. Their descendants are fully *mazehualob* although some exhibit much white admixture and bear Spanish surnames. The priestly leader of the subtribe, a person of great sacredness, has European features and surname, yet the natives claim him as a *mazehual.*

---

[9] "Kaz" is an element with at least two meanings, one referring to that which is bad, and the other to the division of anything.

[10] In Quintana Roo, where almost no Spanish is used, the terms *mestizo* and *vecino* are not employed.

Down in Quintana Roo the tribal group of villagers is in a perpetual state of armed hostility with reference to outsiders, especially the Mexican government. All inside the group must be *mazehualob.* All the outsiders are *dzulob.* For these natives, *dzul* implies no respect; it implies difference and potential enmity. The Quintana Roo natives do not accept the social inferiority which is accepted by the people of Chan Kom. Knowing that the white outsider with whom they come in contact regards *dzul* as a term of respect and *indio* as one of disrespect, the natives of Quintana Roo resent the latter term as applied to them, while toward an outsider with whom they are not yet friendly they will use the word *uinic* ("man") rather than *dzul.* In Chan Kom it is no offense for a *dzul* to speak of a villager as an *indio.* Further the word *xunan* is there used for a woman corresponding to a *dzul,* and similarly implies respect, while in Quintana Roo any married woman, wife of *dzul* or *mazehual,* is a *xunan.*

In short, for all the villagers the essential distinction is that between Us, the in-group, an ethnic unity, and Them, the outsiders, a different ethnic group or groups. For this distinction the principal terms are *dzul* and *mazehual.*[11] In Quintana Roo a *dzul* is not only an alien but an enemy, and a *mazehual* is anyone, even one with evidences of Spanish origin, who lives as a full participant in the society of the village and tribe. In Chan Kom, where the people have accepted status position inside the larger Yucatecan society, *dzul* and *mazehual* take on some of the quality of class terms, and there is admission of the possibility that persons intermediate between *dzul* and *mazehual* not only exist but exist among the participants in village life and culture.

---

[11] In Dzitas *dzul* is sometimes used in the sense of "educated white person not an Indian." In Merida, where much less Maya is spoken, the word is rarely used. It appears in Spanish discourse as a component of depreciative terms or invectives, quoted in the first part of this paper, referring to persons making disapproved claims to superior social position.

The Indian terms *dzul* and *mazehual*, like the Spanish terms *mestizo* and *vecino*, have undergone important changes in meaning. As a society changes, its language changes to fit it. Some glimpses into the history of the words are provided by some documentary materials to which in large part Mr. Roys has kindly called my attention. This line of investigation is here only suggested; it carries the history of ethnic contacts in Yucatan back still farther, even to periods before the coming of the white man.

In the last part of the sixteenth century, if we may trust the Motul dictionary, *dzul* meant "a foreigner from another land." This dictionary further tells us that the word (or a compound) was applied by the Indians to an *encomendero*. The word is a familiar Maya patronymic today, and we read of Indians surnamed Dzul in sixteenth century documents. As a common noun the word must have been in use in Yucatan before the Conquest. Mr. Roys points out that in the Tizimin Chronicle it is said that in Katun 8 Ahau "Mayapan was depopulated by foreigners from the mountains (*uitzil dzul*)." The reference is to the Xiu, who were considered foreigners. *Uitzil* refers to the hills extending southeast from Uxmal. On pages 83–84 of Roys' translation of the Chumayel we read of *dzulob* who seem to be the Itzas, also foreigners to Yucatan. In the Third Chronicle of the Chumayel it is said that "foreigners arrived to eat men." Roys[12] refers to the suggestion of Brinton and Martínez that these invaders were the Caribs, known to be cannibals, but in a personal communication Roys suggests that "chibil" may not mean "to eat" but "to afflict." At any rate these *dzulob* were not Maya. If they were not Mexicans, we have some evidence that the word denoted any foreigner. However, in the Chontal *relación* of Paxbolon, Roys informs me, it is said that

[12] Ralph L. Roys, "The Book of Chilam Balam of Chumayel," p. 142. Carnegie Inst. Wash. Pub. No. 438.

the Chontal-speaking Acalanes had a war with *dzulob.* And the word *dzulob* there occurs in a series of names of tribes (Lacantuns, Ciachob, etc.). The leader of these *dzulob* had a Nahua name (Tzitzimit), and we know that Nahua-speaking people were and are to be found in that part of Tabasco. All these circumstances (with the exception of the possibility that the invaders who came either to eat or to afflict were Caribs) lead to a consideration of the possibility that *dzul* was a Maya word for Mexicans. But whether before the Spaniards arrived the word meant "Mexican" or meant "foreigner," after that event it was applied to the Spaniards, and from connoting chiefly their difference from the Mayas ethnically, came with the Conquest to carry connotations of social superiority until, in certain contexts in Yucatan today, the word "gentleman" (*caballero*) may be given as its equivalent.

The word *mazehual* is Nahua in origin. In Mexico it was used, apparently before the Conquest, to refer to a member of the commoner class, as distinguished from a noble. The Molina dictionary (1571) gives *vasallo* as its Spanish equivalent, and many sixteenth century documents indicate that the *mazehualli* were freemen farmers subordinated to a privileged class. In one such document, for example, quoted in a recent publication,[13] the native upper class, "we *principales*," of Huexotzingo in 1555 petitioned the viceroy, setting forth that from time immemorial the *macehuales* had cultivated land of the *principales*, having none of their own, "gave us chickens and chile and everything for our maintenance, and their wives and children served as in everything we commanded. All this they did so that we might let them sow their crops on our land."

Roys[14] has reviewed the evidence which leads to the conclusion that a similar division of society into two classes, nobles and commoners, existed among the Maya at the time of the Conquest. Roys shows that certain of the ruling families, if not all of them, were regarded as descendants of Nahua invaders of Yucatan. In other words, before the Spaniards came the native Maya society was divided into two classes, and the upper class were thought to be the descendants of invaders or conquerors while the lower class were thought to be of indigenous origin. But members of this upper class were not called *dzulob*, a word reserved as we have seen for outsiders and aliens to the Maya society, or for certain such outsiders. Members of the upper class were called *almehenob*. What the lower class were called at the time of the Conquest is not entirely clear. Roys[15] indicates his opinion that *mazehual* was borrowed from the "Toltec" intruders and hence was in current use before the Spaniards arrived. But occurrences of the word in the earliest Maya texts are not known to the writer. It does not appear in the Motul dictionary. It does appear in Pio Perez' "coordinación" of the words found in Beltrán's grammar. The Book of Chilam Balam of Chumayel, which includes at least one use of the word,[16] is, in the manuscript version we have, much later than the Conquest. So there is negative evidence to support a view that the word *mazehual* was brought into Yucatan as a result of the Spanish invasion.[17] However the question as to the period when the word came into use in Yucatan is to be decided, it is clear that the fundamental pair of terms used in Yucatan since the Conquest by the Indians to distinguish themselves from the Spaniards is made up of an old Maya word for foreigner, or for certain tribes of foreigners, and of a borrowed Nahua word for a member of the lower class of native society.

---

[13] Lesley B. Simpson, The Repartimiento in New Spain and Guatemala. *Ibero-Americana*, 13, *Studies in the Administration of the Indians in New Spain*, p. 21. Berkeley, University of California Press, 1938.

[14] Ralph L. Roys, "The Book of Chilam Balam of Chumayel," pp. 188–195.

This paper, beginning as a summary of terminology used to denote ethnic and status groups recognized in city, town, and village of contemporary Yucatan, has become a tentative outline of the history of changes in ethnic relations in Yucatan. Definitive history is to be done by historians. But in Yucatan old elements of culture and society survive in communities remote from the city. At least it is true that the group relations emphasized in the more remote communities resemble those which must have prevailed in Yucatan when first the Spaniards came in. In Quintana Roo it is the distinction between Us, a culturally and politically independent people, and Them, the threatening invaders. In Chan Kom, as must have been the case in Yucatan as a whole soon after the Conquest, the Indians are still distinguished from the Whites, but intermediate people are recognized and the Indians accept the Whites as a superior status group. Old Dzitas provides a condition which must have first occurred in Yucatan at a time still later: Indians and Spaniards, much intermingled in race and in culture and dwelling together in a well-integrated class society. The later situation in Dzitas and in Merida constitutes a society which it must have taken centuries to develop: ethnic distinctions have almost entirely lapsed; there is a society of Yucatecans, representing one culture, not two, in which class differences depend little on racial factors and in which the individual is free to make his own place.

---

[15] Roys, "The Book of Chilam Balam of Chumayel," p. 188.

[16] *Ibid.*, p. 79, "those who drain (or suck) the poverty-stricken *mazehualob*."

[17] Since the above was written, Mr. Roys has informed me that, in an early sixteenth century Yucatan document, he has found an occurrence of the word *mazehual*.

A comparison of the terms used in these communities, differing as to degree of isolation, aided by a few references to documents, provides a hypothesis as to the history of ethnic and class relations in Yucatan. By way of conclusion the material of this paper may be reviewed in the form of such a hypothesis, but in doing so one must now put first the materials from the remote communities representing the earlier conditions, together with hints from the documents.

Before the Conquest the Maya recognized two social classes. The upper class were regarded as of foreign extraction, the lower class as indigenous. When the Spaniards came in they were denoted *dzulob*, a word already used by the Maya for aliens (or perhaps for certain aliens). The Spaniards then came to supplant the old native privileged class (*almehenob*), and became the upper class in the new society. The Maya continued to speak of them as *dzulob*, this term coming to have a class significance. Either before the Conquest or soon after, the word *mazehualob* was in use for the lower class of the native society; it came to denote the subordinated Indian group in the new society. In the southeast some of the natives never completely accepted their place in the new society, and remained hostile. The War of the Castes reemphasized the hostility between these Maya and the invaders; therefore in that region *dzulob* and *mazehualob* retained or redeveloped connotations emphasizing ethnic differences rather than class relations.

The occupation of the peninsula brought about interbreeding between Spaniards and Indians. Where the offspring was brought up with and entirely identified socially with the Indian group, he might be included with that group and be called *indio* by the Spaniard and *mazehual* by the native. But both groups had to take account of people exhibiting some cultural features of the one group and some features of the other. The Spaniards called such people *mestizos*, the Indians called them *kaz-dzulob.* (It is not clear whether this term was originally depreciative.) Mixed-bloods not identified with the Indian group, especially those bearing Spanish surnames, were given a privileged social position by the Spaniards and were discriminated legally. Because of the different legal position of the *mestizo*, identified by his Spanish surname, he was a full citizen, *vecino*, as the Indian was not; and at least locally this word was used to distinguish such people from Indians.[18] In this way there developed a threefold stratification of the society of Yucatan in general and of Merida in particular. Members of the upper class (itself differentiated into an aristocracy of *encomenderos* and other privileged persons and lesser families, some of whom had Indian blood) were called "Spaniards," "people of the center," or, probably increasingly, *gente de vestido*, from their European dress. Both *mestizos* and Indians wore the folk costume. With the relative increase of *mestizos*, and with the disappearance of both customary and legal distinctions between Indian and *mestizo*, the latter term came to be used, at first loosely and then normally, for all the people who were not the upper class, who were not the *gente de vestido.* These people, taken together, were the lower class of the fundamental two-class system. Thus a *mestizo* became any wearer of the folk costume. By this time, in the city, *indio*, *dzul*, and *mazehual* had become unnecessary terms, while in the rural districts, where assimilation of European blood and customs had gone on more slowly, they still served for the principal social distinctions. Now, as we look at the rapid disappearance of folk dress in the city we foresee another change in the terminology of race and class: *mestizo* will cease, first of all in the city, to denote a class group; it will revert to its original significance of mixed-blood, it will be less used, and the commoner terms will be those having to do with occupation and social privilege. In view of the social and political changes that are taking place in Yucatan as elsewhere, perhaps the underprivileged people of Merida, having been sucsively Indians and mixed-bloods, wearers of the folk costume, and the working people, may yet become "the proletariat." Whatever the change in the society, the language to describe the society will develop accordingly.

---

[18] As *mestizo* came to mean any wearer of the folk costume it was no longer available to distinguish the Spanish-surnamed mixed-bloods from the people with Indian surnames. We may imagine that in conservative backwaters like Dzitas, where the distinction was adhered to, *vecino*, already available as a partial synonym, became *the* term for the local upper class.

THE MAYA NEW EMPIRE

SYLVANUS G. MORLEY
Division of Historical Research
The Carnegie Institution of Washington

THE peninsula of Yucatan, surrounded by water on three sides and separated from the adjacent continental land mass by high mountain ranges, would seem to offer unusually favorable conditions for the study of civilization in a pure state, relatively free from adulteration by alien cultural elements. Here, if anywhere in the New World, because of an almost unique geographic and topographic isolation, a primitive culture may reasonably be expected to have run its course—rise, florescence, and decay, developed and directed by the genius of the indigenous population, conditioned by the environmental factors, but largely free from complicating external influences—an ideal laboratory for the study of human history.

The earlier phases of this story, the beginning and first period of esthetic brilliance, which the writer calls the Maya Old Empire, speaking in a cultural rather than a political sense, developed in the southern half of this region, now the Department of Petén, Guatemala, during the first millennium before Christ. Although these earlier chapters lie beyond the scope of the present contribution, a brief review of Old Empire achievement is essential to a proper understanding of the Maya renaissance in northern Yucatan in the tenth century.

Already as early as the beginning of the Christian era the southern area had been occupied for centuries by an agricultural population which, more probably than not, spoke some early Maya dialect or dialects. It was united by a common religion and, under wise leadership, enjoyed a homogeneous culture well fitted to its needs.

These Old Empire Maya had developed a striking stone architecture. Although the knowledge of metals had not yet come to them they were preeminent in sculpture and the lapidary art. They excelled in ceramics, weaving, feather work, and mosaic. Their knowledge of astronomy surpassed that of the ancient Egyptians, prior to the Ptolemaic period; their hieroglyphic writing was the only original graphic system produced in the Americas; their chronology, exact to the day within a period of some 375,000 years, was as accurate as our own Gregorian calendar;[1] finally, their vigesimal system of numeration, characterized by the concept of zero and numeration by position—developed quite independently of the sexagesimal system of Babylonia and the decimal system of India[2]—is one of the outstanding intellectual attainments of all time.[3]

By the sixth century after Christ, the first quarter of Baktun 9 of the Maya chronological era,[4] all parts of the Old Empire region from the Chiapas and Guatemala highlands in the southwest to Copan, Honduras, in the southeast, the great central area, the Department of Petén and the adjacent Usumacinta Valley to the west, were under intensive occupation and the Maya were already pushing northward into Yucatan.[5]

The new country to the north was as admirably adapted to Maya culture as had been the region to the south. A salubrious climate, its subtropical heat cooled by the never failing trade winds, an excellent, if unusual, water supply, a fertile though sparse soil, which, however, permitted a system of agriculture that, given the rocky character of the country, cannot be bettered even today, a flora unusually rich in food plants, wood in greatest abundance, inexhaustible supplies of easily quarried rock, the native limestone, providing the three essentials for stone architecture: a facile building material, gravel and lime for mortar—all combined to make the northern half of the Yucatan peninsula an exceedingly favorable, indeed congenial, environment for the further and successful continuation of the Maya civilization.

Although the introduction of Maya culture into northern Yucatan as early as the sixth century and its subsequent renaissance there in the tenth century is mainly a problem of archæological research, many other fields of science with their varying techniques have played leading parts in reconstructing the story of man's occupation of this environmentally highly specialized, not to say unique, region.

---

[1] Teeple, 1930, p. 74.

[2] Gandz, 1936, pp. 481–484.

[3] Morley, 1920, p. 28.

[4] The *baktun* was the name of a Maya time period composed of 400 × 360 days, nearly 400 years.

[5] Morley, 1917*a*, p. 145; *idem*, 1920, pp. 457–462; *idem*, 1927; *idem*, 1935, pp. 38–41.

Since the several disciplines involved in making this general survey are too varied to have been found within any single scientific agency, the Division of Historical Research of the Carnegie Institution of Washington, which has been directing this study, correlating its various activities, and largely publishing the results, has had as collaborators not only other departments of the Institution but also a number of outside organizations as well.

Northern Yucatan is a low, flat limestone region with little surface water and no high elevations. Of recent geologic formation (late Tertiary and Quaternary), the land rises from sea level along the north coast to ranges of hills in the south, probably not exceeding 300 meters in height. The principal *sierra* commences in the region of Champoton on the west coast and strikes inland bearing roughly northeast to the modern town of Maxcanu, where, turning sharply southeast, it continues to beyond the towns of Tzuccacab and Peto in southern central Yucatan.

The soil is very shallow with extensive outcroppings of the native rock. In general the vegetation is dense without being high, though the trees increase in size from north to south, reaching the proportions of a tropical forest in the southern and eastern parts of the peninsula.

There is little surface water, no rivers, and only a very few lakes, Bacalar, Chichankanab, and Payegua in southern and central Yucatan being the largest. The water supply of the region consists of large natural openings in the surface, reaching down through solid limestone to the subterranean water table. These vary in depth as the thickness of the surface cap increases, ranging from a few meters in the north to about 50 meters, south of the *sierra*.

These natural wells were called by the Maya *ɔonot*, which their Spanish conquerors softened to *cenote*. They were the chief water supply of the northern peninsula in ancient times. Extensive artificial reservoirs were also constructed and smaller cisterns, called *chultuns*, hollowed out of the living rock to catch the rainfall in sections where *cenotes* are wanting. The *cenotes* of northern Yucatan have been made the subject of a special study, particularly as regards the source of water supply, the composition of the bottoms, and their zoology.[6] The considerable number of *cenotes* in the immediate vicinity of Chichen Itzá, including two unusually large ones at this site, was doubtless a major factor in determining the preeminence of Chichen Itzá[7] in pre-Columbian times.

The climate is healthful, unusually so for the tropics.[8] The extremes of temperature registered at the Carnegie Institution's meteorological station at Chichen Itzá in northeastern Yucatan during the decade from January 27, 1928 to January 26, 1938 ranged from 39°F. on December 15 and 16, 1935 to 107°F. on May 2, 1935. During this ten-year period the mean minimum temperature was 64.4°F., the mean maximum temperature was 90.7°F.[9] Diurnal variations of 30° and 40° are common, especially during the winter months, and this coupled with the well-nigh perfect drainage, provided by the somewhat porous surface limestone, a high percentage of sunshine, and the refreshing trade winds which blow practically continuously throughout the year, are chiefly responsible for the extreme healthfulness of the country, which has been noted by all observers from the sixteenth century to the present day.[10]

The mean annual precipitation at Chichen Itzá for the same ten-year period was 46.43 inches with a variation from 36.09 inches in 1927 to 54.73 inches in 1929.[11] This precipitation is largely concentrated during the rainy season, from May to November, though occasional showers fall throughout the winter and spring months, the dry season. During the winter months heavy storms originating in the north and called locally *nortes*[12] bring lower temperatures, usually in the fifties Fahrenheit, and sometimes cold rains.

The flora of northern Yucatan provides many useful plants.[13]

---

[6] Pearse, Creaser, and Hall, 1936.

[7] The name Chichen Itzá is derived from Maya *chi*, "mouth," *chen*, "wells," and Itzá, the name of the tribe who occupied this site in the tenth century: "the mouths of the wells of the Itzá."

[8] Carnegie Inst. Wash., 1930, p. 110.

[9] The above temperatures are taken from unpublished meteorological records of Chichen Itzá.

[10] Landa, 1928–1929, p. 24; Cogolludo, 1688, p. 176; Carnegie Inst. Wash., 1930, p. 110.

[11] These precipitation figures are taken from unpublished meteorological records of Chichen Itzá.

[12] Shattuck, 1933, pp. 421–422.

[13] Standley, 1930; Lundell, 1937; Bartlett, Standley, Mains, Hedrick, Taylor, Morton, Swallen, Leonard, and Killip, 1935.

Maize heads the list; the cultivation of this grain was without exaggeration the chief occupation of the Maya in ancient times as well as today. It was and still is their principal food and constitutes from 75 to 85 per cent of everything they eat.[14]

Other food plants were beans, squashes, tubers, fruits, chili, chocolate, and many herbs, the last used in flavoring. The medicinal plants are numerous and were extensively employed in ancient times for many specific ailments; indeed a considerable body of aboriginal pharmaceutical lore has survived.[15] Cotton was raised and woven into cloth for garments; henequen, one of the agaves, yielded a fiber from which ropes, cords, and bags were made; and several varieties of gourds provided table utensils and vessels for water, *atole*, *pozole*, and chocolate. Finally, the forest supplied all the materials used in the construction of the thatched houses, the homes of the great majority of the people—the roof timbers, saplings for the sides, vines and withes for binding and tying, and palms for thatching—and also the beams used for the door lintels in the temples and colonnades, the *chico sapote* tree (*Achras zapote* L.).

The fauna of the region[16] supplemented the vegetal food supply, though the Maya are not heavy meat-eaters. Deer were hunted communally for their flesh as well as for their hides, from which sandals were made,[17] while the peccary and other smaller mammals were trapped for their meat. Jaguars were hunted for their pelts, which were worn as cloaks by the priests and nobles, while the brilliant plumage of the abundant bird life[18] was used in making elaborate headdresses and capes. Although the waters along the coast of the peninsula teem with fish, and at some seasons of the year the beaches with turtles, there is little evidence that fish foods played an important part in the general Maya dietary, which, as noted above, was and still is predominantly vegetarian.

Although this region, so admirably adapted for primitive human exploitation, would almost certainly seem to have been occupied by man long before the introduction of Maya culture, probably in the sixth century, few if any traces of this first or Pre-Maya Period have survived.

---

[14] Benedict and Steggerda, 1936, pp. 172, 188.

[15] Roys, 1931.

[16] Gaumer, 1917; Murie, 1935.

[17] Gates, 1937, p. 33, translated from Landa, 1928–1929, p. 150.

[18] Van Tyne, 1935.

The stone architecture with its typical corbel-arched roofs, façade decoration, and specialized ground plans is essentially Maya and stems directly from the stone architecture of the Maya Old Empire to the south, though naturally showing regional modifications.

The hieroglyphic writing of the northern sites derives without break from the inscriptions of the southern sites, and the chronologic sequence of the archæological remains of the two regions is established beyond all doubt by the priority of the dated monuments of the south.

The sculpture of northern Yucatan, though far inferior to that of most of the southern sites, is, nevertheless, a lineal descendant thereof; its subject matter is similar and its technique identical. In the north, however, sculpture as an independent art languished, being more and more restricted to the embellishment of architecture.

Indeed, were it not for the highly individual character of the ceramic wares of northern Yucatan, largely if not entirely of indigenous origin, one might almost hazard the conjecture that the bearers of Maya culture to this region in the sixth century were the first people to occupy and colonize it.

Excavation in the caves of northern Yucatan, which are numerous and, in some cases, of vast extent, has established the fact of their former occupancy by man, but it is significant to observe that the types of pottery and the bone and stone implements found in them in no way differ from the corresponding artifacts found associated with the surface archæological remains.[19] This cave material indicates a culture not only contemporaneous with but also identical with that of the surface sites, which we have seen, in so far as their architecture and hieroglyphic inscriptions are concerned, is Old Empire Maya and was derived directly from southern sources.

It is the ceramic material, however, found both in the caves and at the surface sites, which seems to hold the key to the problem of man's occupation of northern Yucatan prior to the advent of Maya culture.

The ceramic wares of northern Yucatan, as established by considerable, though perhaps as yet insufficient excavation to permit certain determination, are in all probability not derived from contemporaneous Old Empire wares and owe little or nothing to outside inspiration.

---

[19] Mercer, 1896. See Pearse and collaborators, 1938, for a study of the fauna of the caves of Yucatan.

Although much work yet remains to be done both in the caves and in the kitchen middens along the coast of the peninsula before final conclusions may be reached, present indications are that the ceramic wares of northern Yucatan, generally speaking, are of indigenous origin.

But we have seen that the Maya culture of northern Yucatan—the typical architecture and unique hieroglyphic writing—was surely introduced directly from the Old Empire cities in the south, and if this be true, why were the contemporary Old Empire ceramic types not introduced at the same time?

The best, if not indeed the only, explanation of the observed archæological facts seems to be that the northern half of the peninsula was occupied, presumably by an agricultural people, before the advent of Maya culture from the south, and further that these previous occupants of the region had already developed their own ceramic wares, which they retained while adopting at the same time the stone architecture and hieroglyphic writing of the newcomers with their accompanying astronomic learning and religious philosophy.

In the present state of knowledge it is idle to speculate, except very generally, as to the first inhabitants of the northern half of the peninsula.

It has been suggested that they were agriculturists because agriculture, especially as applied to the cultivation of maize, must have preceded by centuries or even millenniums the development of Maya culture in the older southern centers. There is no reason to suppose that these early inhabitants did not also cultivate maize; indeed, to eke out an existence in the waterless forests of northern Yucatan without agriculture would have been little short of impossible.

It is probably to be assumed also that they were some branch of the Maya stock, not only because they assimilated the new cultural complex from the south so readily, but, perhaps more significant, because many of the modern Maya of northern Yucatan could well have served as the original models for the Old Empire reliefs. Their physical type is so unusual—they are one of the world's most brachycephalic people[20]—and their resemblance to the figures on the Old Empire monuments is so striking (plate 1) that they could only have been of the same blood as the people who first brought Maya culture to northern Yucatan.

[20] Williams, 1931, pp. 104–108; Steggerda, 1932, pp. 61–66, 83–85, 97, 99.

If we assume that the original inhabitants were a Maya-speaking people, practicing maize cultivation, but without enjoying those higher manifestations of the culture now known as Maya, we are probably not far from correct in the general picture of the people who lived in the northern half of the Yucatan Peninsula during this first or Pre-Maya Period (table 1, page 542).

Maya culture, as defined above, seems to have first reached northern Yucatan by way of the east coast of the peninsula. The archæological, documentary, and traditional evidence supporting this fact is clear and unmistakable.

The earliest dated monument now known in the northern half of the peninsula is Stela 1 at Tulum on the east coast, presenting the early Initial Series 9.6.10.0.0 (table 1). This may be said to mark the beginning of the second or Maya Period of Yucatan history.[21]

But happily Tulum is not the only link in the chronologic chain connecting the Old and New Empires along the eastern side of the peninsula. Stela 1 at Ichpaatun, located near the western shore of Chetumal Bay, some 140 miles south-southwest of Tulum, records an Initial Series only thirty years later, 9.8.0.0.0;[22] while the very large site of Coba in northeastern Yucatan, about 40 miles northwest of Tulum, has a number of inscribed stelæ recording dates between 9.9.0.0.0 and 9.12.10.0.0, the earliest being only fifty years later than the date of the Tulum monument.[23]

Here is irrefutable archæological evidence that very early in recorded Maya history, i.e. during the last century of the Early Period of the Old Empire, A.D. 534–633 (table 1), a chain of sites reached northward from the northeastern corner of the Old Empire region, along the eastern coast of the peninsula, and inland to Coba in northeastern Yucatan.

In the *u kahlay katunob* of the Books of Chilam Balam of Mani and Tizimin,[24] it is clearly stated that Chichen Itzá, only 50 miles west and slightly north of Coba, was discovered while the Maya were in the Province of Ziyancaan Bakhalal.[25]

---

[21] Howe, 1911; Carnegie Inst. Wash., 1916, pp. 338–339; Morley, 1917; Lothrop, 1924.

[22] Carnegie Inst. Wash., 1925, pp. 260, 277; Gann, 1926, pp. 48–55, 248.

[23] Carnegie Inst. Wash., 1926, pp. 274–277; Morley, 1927.

[24] The so-called Books of Chilam Balam are sixteenth, seventeenth, and eighteenth century redactions in Spanish script of earlier manuscripts in the Maya hieroglyphic writing. The chronicles in these manuscripts were called *u kahlay katunob* or "count of the katuns," the katun being a period of about 20 years, i.e. 20 × 360 days. These have been translated by several different writers: Brinton, 1882, Martínez Hernández, 1927, and Roys, 1933.

Brinton has pointed out that Ziyancaan, meaning in Maya *ziyan*, "birth," and *caan*, "sky" or "heavens," is probably a picturesque allusion to the seashore as the place where the sky is born,[26] and since Bakhalal, only slightly Hispanicized in form as Bacalar, still survives in the name of the largest lake in southeastern Yucatan, the documentary evidence entirely substantiates the archæological evidence to the effect that Maya culture first entered Yucatan from the southeast.

Finally, a very early tradition preserved by Father Bernardo de Lizana during the second decade of the seventeenth century tells the same story:

> The early fathers who first emplanted the Faith of Christ in Yucatan knew that the people here came part from the west and part from the east, and thus in their ancient language they call the east in another way than today. Today they call the east LiKin which is the same as where the sun rises, as with us. And the west they call ChiKin which is the same as the fall or end of the sun or where it hides itself, as with us. And anciently they called the east, Cenial and the west Nohenial, cenial which is to say the small descent: and nohenial the great descent. And the reason they say this is because from the part of the east there came down to this land few people, and by that of the west, many and with this syllable they understand few or many, the east and the west; and the few people from one part and the many from the other.[27]

Thus the traditional evidence, so far as it goes, also agrees with the archæological indications and the documentary statements to the effect that the first entrance of Maya culture in northern Yucatan was along the east coast, i.e. from the east or southeast.

As to when this took place both the *u kahlay katunob* and the inscriptions are again in general agreement.

The *u kahlay katunob* from the Book of Chilam Balam of Mani states that the discovery of the Province of Ziyancaan Bakhalal took place in a Katun 6 Ahau, 692–711, and that during the sixty years the Maya were at Bakhalal (711–771) occurred the discovery of Chichen Itzá.[28] The *u kahlay katunob* from the Book of Chilam Balam of Tizimin states that Chichen Itzá was heard of and the

---

[25] Brinton, 1882, pp. 95, 96, 100–101, 138, 144; Martínez Hernández, 1927, pp. 7, 15.

[26] Brinton, 1882, p. 124.

[27] Lizana, 1893, ffº *3v 4r*.

[28] Brinton, 1882, pp. 95–96, 100–101; Martínez Hernández, 1927, p. 7.

TABLE 1. THE MAYA NEW EMPIRE

Corresponding Old Empire period	New Empire period	Long Count dates (the Initial Series)	*U kahlay katunob*	Gregorian equivalents Anno Domini	Events and dated monuments
Early Period	First or Pre-Maya Period	9. 0. 0.0.0 8 Ahau 13 Ceh....	**8 Ahau**	435, December 11	
		9. 1. 0.0.0 6 Ahau 13 Yaxkin.	6 Ahau	455, August 28	
		9. 2. 0.0.0 4 Ahau 13 Uo.....	4 Ahau	475, May 15	
		9. 3. 0.0.0 2 Ahau 18 Muan..	2 Ahau	495, January 30	
		9. 4. 0.0.0 13 Ahau 18 Yax...	13 Ahau	514, October 18	
		9. 5. 0.0.0 11 Ahau 18 Tzec..	11 Ahau	534, July 5	
		9. 6. 0.0.0 9 Ahau 3 Uayeb...	9 Ahau	554, March 22	
		9. 6.10.0.0 8 Ahau 13 Pax....		564, January 29	Tulum, Stela 1 (9.6.10.0.0)
		9. 7. 0.0.0 7 Ahau 3 Kankin..	7 Ahau	573, December 7	
		9. 8. 0.0.0 5 Ahau 3 Chen. ...	5 Ahau	593, August 24	Ichpaatun, Stela 1 (9.8.0.0.0)
		9. 9. 0.0.0 3 Ahau 3 Zotz.....	3 Ahau	613, May 12	Coba, Stelæ 1–6 (9.9.0.0.0–9.12.10.0.0)
		9.10. 0.0.0 1 Ahau 8 Kayab...	1 Ahau	633, January 27	
Middle Period		9.11. 0.0.0 12 Ahau 8 Ceh....	12 Ahau	652, October 14	Island of Jaina, Stela 1 (9.11.0.0.0)
		9.12. 0.0.0 10 Ahau 8 Yaxkin.	10 Ahau	672, July 1	
		9.12.10.0.0 9 Ahau 18 Zotz...		682, May 10	
		9.13. 0.0.0 8 Ahau 8 Uo......	**8 Ahau**	692, March 18	Chichen Itzá first heard of
		9.14. 0.0.0 6 Ahau 13 Muan..	6 Ahau	711, December 5	Chichen Itzá occupied by Old Empire Maya
		9.15. 0.0.0 4 Ahau 13 Yax....	4 Ahau	731, August 22	Etzna, Stela 2 (9.15.0.0.0)
Great Period	Second or Maya Period	9.16. 0.0.0 2 Ahau 13 Tzec...	2 Ahau	751, May 9	Santa Rosa Xtampak, Stelæ 5 and 7 (9.16.0.0.0)
		9.16.13.0.0 2 Ahau 8 Uo......		764, March 1	Holactun, Temple of the Initial Series (9.16.13.0.0)
		9.17. 0.0.0 13 Ahau 18 Cumhu	13 Ahau	771, January 24	Calendar put in order
		9.18. 0.0.0 11 Ahau 18 Mac. .	11 Ahau	790, October 11	
		9.19. 0.0.0 9 Ahau 18 Mol ...	9 Ahau	810, June 28	
		10. 0. 0.0.0 7 Ahau 18 Zip....	7 Ahau	830, March 15	
		10. 1. 0.0.0 5 Ahau 3 Kayab...	5 Ahau	849, November 30	
		10. 2. 0.0.0 3 Ahau 3 Ceh.....	3 Ahau	869, August 17	
		10. 2.10.0.0 2 Ahau 13 Chen...		879, June 26	Chichen Itzá lintel (10.2.10.0.0)
		10. 3. 0.0.0 1 Ahau 3 Yaxkin..	1 Ahau	889, May 4	Closing date on Old Empire stelæ: Uaxactun, Stela 12; Xultun, Stela 10; Xamantun, Stela 1 (10.3.0.0.0)
		10. 4. 0.0.0 12 Ahau 3 Uo.....	12 Ahau	909, January 20	Tzibanche, jadeite gorget (10.4.0.0.0)
		10. 5. 0.0.0 10 Ahau 8 Muan..	10 Ahau	928, October 7	
		10. 6. 0.0.0 8 Ahau 8 Yax.....	**8 Ahau**	948, June 24	Chakanputun abandoned. The Itzá enter northern Yucatan
		10. 7. 0.0.0 6 Ahau 8 Tzec	6 Ahau	968, March 11	
		10. 8. 0.0.0 4 Ahau 13 Cumhu	4 Ahau	987, November 27	[illegible]

THIRD OR MEXICAN (ITZÁ) PERIOD	The League of Mayapan	10.10. 0.0.0 13 Ahau 13 Mol...	13 Ahau	1027, May 2	League of Mayapan begins
		10.11. 0.0.0 11 Ahau 13 Zip...	11 Ahau	1047, January 17	
		10.12. 0.0.0 9 Ahau 18 Pax....	9 Ahau	1066, October 4	
		10.13. 0.0.0 7 Ahau 18 Zac....	7 Ahau	1086, June 21	
		10.14. 0.0.0 5 Ahau 18 Xul....	5 Ahau	1106, March 9	
		10.15. 0.0.0 3 Ahau 18 Pop....	3 Ahau	1125, November 24	
		10.16. 0.0.0 1 Ahau 3 Muan...	1 Ahau	1145, August 11	
		10.17. 0.0.0 12 Ahau 3 Yax....	12 Ahau	1165, April 28	
		10.18. 0.0.0 10 Ahau 3 Tzec...	10 Ahau	1185, January 13	
		10.18.10.0.0 9 Ahau 13 Uo.....		1194, November 22	Plot of Hunnac Ceel. Itzá "driven from Chichen Itzá" by the Cocom. End of League of Mayapan
		10.19. 0.0.0 8 Ahau 8 Cumhu..	**8 Ahau**	1204, September 30	
	The Hegemony of Mayapan	11. 0. 0.0.0 6 Ahau 8 Mac. ...	6 Ahau	1224, June 17	
		11. 1. 0.0.0 4 Ahau 8 Mol.....	4 Ahau	1244, March 4	
		11. 2. 0.0.0 2 Ahau 8 Zip......	2 Ahau	1263, November 20	
		11. 3. 0.0.0 13 Ahau 13 Pax...	13 Ahau	1283, August 7	
		11. 4. 0.0.0 11 Ahau 13 Zac...	11 Ahau	1303, April 25	
		11. 5. 0.0.0 9 Ahau 13 Xul....	9 Ahau	1323, January 10	
		11. 6. 0.0.0 7 Ahau 13 Pop....	7 Ahau	1342, September 27	
		11. 7. 0.0.0 5 Ahau 18 Kankin.	5 Ahau	1362, June 14	
		11. 8. 0.0.0 3 Ahau 18 Chen...	3 Ahau	1382, March 1	
		11. 9. 0.0.0 1 Ahau 18 Zotz....	1 Ahau	1401, November 17	
		11.10. 0.0.0 12 Ahau 3 Cumhu.	12 Ahau	1421, August 4	
		11.11. 0.0.0 10 Ahau 3 Mac....	10 Ahau	1441, April 21	Destruction of Mayapan by Xiu and other Maya chieftains; all large cities abandoned
FOURTH OR PERIOD OF DECADENCE		11.12. 0.0.0 8 Ahau 3 Mol.....	**8 Ahau**	1461, January 6	
		11.13. 0.0.0 6 Ahau 3 Zip......	6 Ahau	1480, September 23	
		11.14. 0.0.0 4 Ahau 8 Pax.....	4 Ahau	1500, June 11	
		11.15. 0.0.0 2 Ahau 8 Zac.....	2 Ahau	1520, February 27	
		11.16. 0.0.0 13 Ahau 8 Xul....	13 Ahau	1539, November 14	
		11.17. 0.0.0 11 Ahau 8 Pop....	11 Ahau	1559, August 1	
		11.18. 0.0.0 9 Ahau 13 Kankin	9 Ahau	1579, April 18	
		11.19. 0.0.0 7 Ahau 13 Chen...	7 Ahau	1599, January 3	
		12. 0. 0.0.0 5 Ahau 13 Zotz....	5 Ahau	1618, September 20	
		12. 1. 0.0.0 3 Ahau 18 Kayab.	3 Ahau	1638, June 7	
		12. 2. 0.0.0 1 Ahau 18 Ceh....	1 Ahau	1658, February 22	
		12. 3. 0.0.0 12 Ahau 18 Yaxkin	12 Ahau	1677, November 9	
		12. 4. 0.0.0 10 Ahau 18 Uo....	10 Ahau	1697, July 27	Tayasal destroyed by Martín de Ursúa y Arizmendi 136 days before the beginning of Katun 8 Ahau
		12. 5. 0.0.0 8 Ahau 3 Pax.....	**8 Ahau**	1717, April 14	

TABLE 1. THE MAYA NEW EMPIRE

Corresponding Old Empire period	New Empire period		Long Count dates (the Initial Series)	*U kahlay katunob*	Gregorian equivalents Anno Domini	Events and dated monuments
Early Period	First or Pre-Maya Period		9. 0. 0.0.0 8 Ahau 13 Ceh	**8 Ahau**	435, December 11	
			9. 1. 0.0.0 6 Ahau 13 Yaxkin	6 Ahau	455, August 28	
			9. 2. 0.0.0 4 Ahau 13 Uo	4 Ahau	475, May 15	
			9. 3. 0.0.0 2 Ahau 18 Muan	2 Ahau	495, January 30	
			9. 4. 0.0.0 13 Ahau 18 Yax	13 Ahau	514, October 18	
			9. 5. 0.0.0 11 Ahau 18 Tzec	11 Ahau	534, July 5	
			9. 6. 0.0.0 9 Ahau 3 Uayeb	9 Ahau	554, March 22	
			9. 6.10.0.0 8 Ahau 13 Pax		564, January 29	Tulum, Stela 1 (9.6.10.0.0)
	Second or Maya Period		9. 7. 0.0.0 7 Ahau 3 Kankin	7 Ahau	573, December 7	
			9. 8. 0.0.0 5 Ahau 3 Chen	5 Ahau	593, August 24	Ichpaatun, Stela 1 (9.8.0.0.0)
			9. 9. 0.0.0 3 Ahau 3 Zotz	3 Ahau	613, May 12	Coba, Stelæ 1–6 (9.9.0.0.0–9.12.10.0.0)
			9.10. 0.0.0 1 Ahau 8 Kayab	1 Ahau	633, January 27	
Middle Period			9.11. 0.0.0 12 Ahau 8 Ceh	12 Ahau	652, October 14	Island of Jaina, Stela 1 (9.11.0.0.0)
			9.12. 0.0.0 10 Ahau 8 Yaxkin	10 Ahau	672, July 1	
			9.12.10.0.0 9 Ahau 18 Zotz		682, May 10	
			9.13. 0.0.0 8 Ahau 8 Uo	**8 Ahau**	692, March 18	Chichen Itzá first heard of
			9.14. 0.0.0 6 Ahau 13 Muan	6 Ahau	711, December 5	Chichen Itzá occupied by Old Empire Maya
			9.15. 0.0.0 4 Ahau 13 Yax	4 Ahau	731, August 22	Etzna, Stela 2 (9.15.0.0.0)
Great Period			9.16. 0.0.0 2 Ahau 13 Tzec	2 Ahau	751, May 9	Santa Rosa Xtampak, Stelæ 5 and 7 (9.16.0.0.0)
			9.16.13.0.0 2 Ahau 8 Uo		764, March 1	Holactun, Temple of the Initial Series (9.16.13.0.0)
			9.17. 0.0.0 13 Ahau 18 Cumhu	13 Ahau	771, January 24	Calendar put in order
			9.18. 0.0.0 11 Ahau 18 Mac	11 Ahau	790, October 11	
			9.19. 0.0.0 9 Ahau 18 Mol	9 Ahau	810, June 28	
			10. 0. 0.0.0 7 Ahau 18 Zip	7 Ahau	830, March 15	
			10. 1. 0.0.0 5 Ahau 3 Kayab	5 Ahau	849, November 30	
			10. 2. 0.0.0 3 Ahau 3 Ceh	3 Ahau	869, August 17	
			10. 2.10.0.0 2 Ahau 13 Chen		879, June 26	Chichen Itzá lintel (10.2.10.0.0)
			10. 3. 0.0.0 1 Ahau 3 Yaxkin	1 Ahau	889, May 4	Closing date on Old Empire stelæ: Uaxactun, Stela 12; Xultun, Stela 10; Xamantun, Stela 1 (10.3.0.0.0)
			10. 4. 0.0.0 12 Ahau 3 Uo	12 Ahau	909, January 20	Tzibanche, jadeite gorget (10.4.0.0.0)
			10. 5. 0.0.0 10 Ahau 8 Muan	10 Ahau	928, October 7	
			10. 6. 0.0.0 8 Ahau 8 Yax	**8 Ahau**	948, June 24	Chakanputun abandoned. The Itzá enter northern Yucatan
	Third or Mexican (Itzá) Period		10. 7. 0.0.0 6 Ahau 8 Tzec	6 Ahau	968, March 11	
			10. 8. 0.0.0 4 Ahau 13 Cumhu	4 Ahau	987, November 27	Chichen Itzá occupied by the Itzá. Mayapan founded by Kukulcan
		The League of Mayapan	10. 9. 0.0.0 2 Ahau 13 Mac	2 Ahau	1007, August 15	Uxmal founded by the Xiu. The League of Mayapan begins
			10.10. 0.0.0 13 Ahau 13 Mol	13 Ahau	1027, May 2	
			10.11. 0.0.0 11 Ahau 13 Zip	11 Ahau	1047, January 17	
			10.12. 0.0.0 9 Ahau 18 Pax	9 Ahau	1066, October 4	
			10.13. 0.0.0 7 Ahau 18 Zac	7 Ahau	1086, June 21	
			10.14. 0.0.0 5 Ahau 18 Xul	5 Ahau	1106, March 9	
			10.15. 0.0.0 3 Ahau 18 Pop	3 Ahau	1125, November 24	
			10.16. 0.0.0 1 Ahau 3 Muan	1 Ahau	1145, August 11	
			10.17. 0.0.0 12 Ahau 3 Yax	12 Ahau	1165, April 28	
			10.18. 0.0.0 10 Ahau 3 Tzec	10 Ahau	1185, January 13	
			10.18.10.0.0 9 Ahau 13 Uo		1194, November 22	Plot of Hunnac Ceel. Itzá "driven from Chichen Itzá" by the Cocom. End of League of Mayapan
			10.19. 0.0.0 8 Ahau 8 Cumhu	**8 Ahau**	1204, September 30	
		The Hegemony of Mayapan	11. 0. 0.0.0 6 Ahau 8 Mac	6 Ahau	1224, June 17	
			11. 1. 0.0.0 4 Ahau 8 Mol	4 Ahau	1244, March 4	
			11. 2. 0.0.0 2 Ahau 8 Zip	2 Ahau	1263, November 20	
			11. 3. 0.0.0 13 Ahau 13 Pax	13 Ahau	1283, August 7	
			11. 4. 0.0.0 11 Ahau 13 Zac	11 Ahau	1303, April 25	
			11. 5. 0.0.0 9 Ahau 13 Xul	9 Ahau	1323, January 10	
			11. 6. 0.0.0 7 Ahau 13 Pop	7 Ahau	1342, September 27	
			11. 7. 0.0.0 5 Ahau 18 Kankin	5 Ahau	1362, June 14	
			11. 8. 0.0.0 3 Ahau 18 Chen	3 Ahau	1382, March 1	
			11. 9. 0.0.0 1 Ahau 18 Zotz	1 Ahau	1401, November 17	
			11.10. 0.0.0 12 Ahau 3 Cumhu	12 Ahau	1421, August 4	
			11.11. 0.0.0 10 Ahau 3 Mac	10 Ahau	1441, April 21	Destruction of Mayapan by Xiu and other Maya chieftains; all large cities abandoned
	Fourth or Period of Decadence		11.12. 0.0.0 8 Ahau 3 Mol	**8 Ahau**	1461, January 6	
			11.13. 0.0.0 6 Ahau 3 Zip	6 Ahau	1480, September 23	
			11.14. 0.0.0 4 Ahau 8 Pax	4 Ahau	1500, June 11	
			11.15. 0.0.0 2 Ahau 8 Zac	2 Ahau	1520, February 27	
			11.16. 0.0.0 13 Ahau 8 Xul	13 Ahau	1539, November 14	
			11.17. 0.0.0 11 Ahau 8 Pop	11 Ahau	1559, August 1	
			11.18. 0.0.0 9 Ahau 13 Kankin	9 Ahau	1579, April 18	
			11.19. 0.0.0 7 Ahau 13 Chen	7 Ahau	1599, January 3	
			12. 0. 0.0.0 5 Ahau 13 Zotz	5 Ahau	1618, September 20	
			12. 1. 0.0.0 3 Ahau 18 Kayab	3 Ahau	1638, June 7	
			12. 2. 0.0.0 1 Ahau 18 Ceh	1 Ahau	1658, February 22	
			12. 3. 0.0.0 12 Ahau 18 Yaxkin	12 Ahau	1677, November 9	
			12. 4. 0.0.0 10 Ahau 18 Uo	10 Ahau	1697, July 27	Tayasal destroyed by Martín de Ursúa y Arizmendi 136 days before the beginning of Katun 8 Ahau
			12. 5. 0.0.0 8 Ahau 3 Pax	**8 Ahau**	1717, April 14	

discovery of the Province of Ziyancaan took place in a Katun 8 Ahau, 672–692.[29] Finally, the first *u kahlay katunob* from the Book of Chilam Balam of Chumayel gives the date of the discovery of Chichen Itzá as Katun 6 Ahau, 692–711, again.[30]

The dates of the Tulum, Ichpaatun, and Coba stelæ in the Gregorian calendar (table 1) are 564, 593, and 613–682 respectively. These indicate that a chain of dated Maya sites extended up the east coast of the peninsula and as far inland as Coba from 564 to 682. But we have just seen that the Province of Ziyancaan Bakhalal of the *u kahlay katunob* is probably to be identified with the general east coast region. If so, the native chronicles indicate that this province was settled by the Maya sometime during the century and a half between 613 and 771, probably nearer the former than the latter date, and that while they were in this section, the largest center of which seems to have been Coba,[31] the discovery of Chichen Itzá took place. This much is definitely established both by the archæological and by the documentary evidence.

Coba would seem to have been the largest Old Empire center in northern Yucatan. The range of its surely dated monuments, roughly seventy years—613–682—strongly suggests that it may even have been the site from which Chichen Itzá was "first heard of" between 672 and 771; and other archæological evidence is not wanting in support of the same conclusion.

From Coba westward there stretches a stone causeway 62½ miles long to the ruins of Yaxuna, about 12 miles southwest of Chichen Itzá. This causeway bears generally west with several minor changes of direction, all but the last of which, however, are within 20 miles of Coba.[32] This would appear to indicate that the construction of this causeway was started at its eastern or Coba end, that it was built from the east toward the west, and that its several changes of direction during the first 20 miles from Coba were for the purpose of passing existing settlements; and, finally, that the relative straightness of the last (western) two-thirds is due to the fact that there were fewer Maya settlements to the west when the causeway was built.

[29] Brinton, 1882, pp. 138, 144; Martínez Hernández, 1927, p. 15.

[30] Brinton, 1882, pp. 153, 158; Martínez Hernández, 1927, p. 24; Roys, 1933, pp. 47, 135.

[31] Thompson, Pollock, and Charlot, 1932.

[32] Villa, 1934.

Weighing all the available evidence, archæological, documentary, and traditional, it seems highly probable that Maya culture first reached northern Yucatan from the south by way of the east coast during the sixth century of the Christian era, i.e. just before or just after the end of the first quarter of Baktun 9 of the Maya era. During this quarter of Baktun 9, 435–534, as the writer has shown elsewhere,[33] Maya culture first spread from the original cultural focus in northeastern Petén (Uaxactun) to the peripheral regions lying to the southeast (Copan), southwest (Chiapas highlands), west (Usumacinta Valley), and northeast (east coast of Yucatan).

By the end of the Early Period of the Old Empire, 9.10.0.0.0, 633 (table 1), this eastern wave of Maya culture, "the little descent," had subsided and "the great descent" from the west, i.e. the southwest, was getting under way.

Although the archæological evidence is not so satisfactory for the west coast region—the writer believes largely because the southwestern parts of the Campeche are, archæologically speaking, less well known than the east coast region—corroboratory items are not wanting even here. There are at least three sites in western or central Campeche which have Initial Series monuments: Stela 1 on the Island of Jaina, adjoining the west coast of the peninsula, recording the date 9.11.0.0.0, 652;[34] Stela 2 at Etzna, recording the date 9.15.0.0.0, 731;[35] and the Temple of the Initial Series at Holactun, probably recording the date 9.16.13.0.0, 764,[36] the last two sites being in central Campeche.

Though fewer, the dated monuments in the western part of the peninsula indicate none the less surely that Maya culture had penetrated this region about the beginning of the Middle Period of the Old Empire, 633 (Stela 1, Island of Jaina, 652, table 1) and at least two other sites, Etzna and Holactun, in this region were erecting sculptured stone monuments by the beginning of the Great Period, 731 and 764 respectively. This, the writer believes, was the beginning of the *nohenial* or "great descent" mentioned in the tradition preserved by Father Lizana previously quoted (page 541).

The explorations of the Fifteenth, Sixteenth, and Seventeenth

---

[33] Morley, 1917*a*, p. 145; *idem*, 1920, pp. 457–462; *idem*, 1927; *idem*, 1935, pp. 38–41; *idem*, 1937–1938, vol. IV, pp. 323–326.

[34] Carnegie Inst. Wash., 1924, p. 211.

[35] *Idem*, 1927, p. 263.

[36] Thompson, 1929, p. 226; *idem*, 1931, pp. 354–356.

Central American Expeditions of the Carnegie Institution in 1932,[37] 1933,[38] and 1934[39] respectively, in southern Campeche, have established that as early as the end of the Early Period, Maya culture was pushing northward into Yucatan along a third line of migration about midway between the east and west coast routes already discussed. Uxul and the very large site of Calakmul in southern Campeche were found to have dates as early as 613 (9.9.0.0.0)[40] and 623 (9.9.10.0.0)[41] respectively, perhaps about the same time.

Although the architecture of southern Yucatan and northern Campeche has not yet been studied sufficiently to permit final conclusions, it appears, nevertheless, that the architecture of the former, the so-called Puc or hilly region, may have been produced by the Maya who entered the peninsula from the southwest—the first waves of "the great descent"; while the architecture of northern and central Campeche, the Chenes region, would seem to have been built by the Maya who came from the south, entering the northern half of the peninsula by the third or intermediate line of migration, just mentioned, perhaps about the same time.

During the Great Period of the Old Empire, 731–909 (table 1), the Maya centers that had grown up in the north, the writer's New Empire, were peripheral to the great southern cities, which were the chief foci of cultural inspiration, and throughout the Great Period even the largest of the northern sites remained provincial.

Although we have seen that Chichen Itzá was first heard of as early as 672–692, the earliest surely dated sculpture at this site is almost two centuries later—the re-used lintel from the Temple of the Initial Series dedicated in 879 (table 1).[42] Coba was much larger and more important than Chichen Itzá at this time, indeed the Class 3 site of Yaxuna, the western terminus of the great Coba-Yaxuna causeway, would seem to have been larger than Chichen Itzá in the eighth century, since this causeway terminates at Yaxuna instead of Chichen Itzá only 12 miles distant. Although the point cannot be established without considerable excavation at Yaxuna, the writer ventures the prediction that Maya culture was first established at Chichen Itzá directly from Yaxuna.

---

[37] Carnegie Inst. Wash., 1932, pp. 95–96.

[38] *Idem*, 1933, pp. 89–92.

[39] *Idem*, 1934, pp. 93–95.

[40] *Ibid.*, p. 95.

[41] *Idem*, 1932, p. 96.

[42] Morley, 1913; *idem*, 1915, pp. 197, 199; *idem*, 1920, pp. 393, 458, 527, 531, 572, 573, 586.

During the last half of the ninth century and the early years of the tenth century, the fortunes of the Old Empire were fast ebbing. The writer has reviewed elsewhere[43] the several causes which have been variously suggested as having been responsible for the decay and final collapse of the Maya Old Empire and these need not be repeated here other than to state that economic failure, inability to make the land produce sufficient maize under the Maya system of agriculture[44] to support the increasing population, would seem to have been the chief factor responsible for the decay of the Old Empire and the subsequent abandonment of its different centers.

In 889 (10.3.0.0.0) the last sculptured monuments were being dedicated in the Old Empire: Stela 12 at Uaxactun,[45] Stela 10 at Xultun,[46] both in northern Petén, and Stela 1 at Xamantun in southern Campeche;[47] and twenty years later, 909 (10.4.0.0.0), the last date in the south was being engraved on a jadeite gorget from Tzibanche in southern Quintana Roo.[48]

The writer believes that not long after the close of monumental activity in the southern cities, fifty or a hundred years at most, they were finally abandoned as religious and ceremonial centers and the population moved elsewhere, some northward into the New Empire region and others southward into the foothills of the Guatemalan cordillera.

Other events were making in the tenth century that were destined to carry forward the Maya of the New Empire to a true renaissance during which architecture was to flourish as never before and their civilization to flower for the second and last time.

In Katun 8 Ahau, 928–948 (table 1), a tribe of Maya-speaking people called the Itzá, who had been living in the southwestern part of the peninsula in and around Chakanputun (the modern Champoton, south of Campeche) for some two to two and a half centuries, began moving slowly northeastward across the peninsula

[43] *Idem*, 1920, pp. 442–462; *idem*, 1935, pp. 34–41.

[44] Redfield and Villa, 1934, pp. 42–47.

[45] Carnegie Inst. Wash., 1916, pp. 339–340.

[46] *Idem*, 1920, pp. 322, 324.

[47] *Idem*, 1932, p. 96; *idem*, 1933, p. 89.

[48] Gann, 1935, pp. 161–163.

and established themselves at Chichen Itzá about forty years later in a Katun 4 Ahau, 968–987 (table 1). The three principal *u kahlay katunob* are in agreement as to the katun in which the Itzá abandoned Chakanputun and a fourth—the third chronicle from the Book of Chilam Balam of Chumayel[49]—is a shade more explicit, stating that Chakanputun perished by fire "which destroyed it quickly and suddenly consumed it,"[50] and that this event took place in the seventh tun of Katun 8 Ahau, i.e. in 935.

The abandonment of Chakanputun the writer has chosen as marking the beginning of the third or Mexican Period of Yucatan history (table 1).

We are entering now upon firmer historical ground. Although the monuments dated by means of the infallible Initial Series have ceased, a number of sculptures dated by the less accurate Period Ending dates occasionally supplement the record, while the *u kahlay katunob* still remain as our chief sources for the reconstruction of early Yucatan history. Finally, a new source appears in the Spanish chroniclers of the sixteenth and seventeenth centuries. Of these, Bishop Diego de Landa, writing in 1566, less than a generation after the Spanish Conquest, is the most complete and credible.

Says Landa in describing these later waves of migration, which the writer believes were a continuation of "the great descent" from the southwest that had begun three centuries earlier:

> Chichén Itzá is finely situated ten leagues from Izamal and eleven from Valladolid, and they tell that it was ruled by three lords, brothers that came to the country from the West. These were very devout, built very handsome temples, and lived unmarried and most honorably. One of them either died or went away, whereupon the others conducted themselves unjustly and wantonly, for which they were put to death. . . .
>
> The opinion of the Indians is that with the Itzás who settled Chichén Itzá there ruled a great lord named Cuculcán, as an evidence of which the principal building [the Castillo] is called Cuculcán.
>
> They say that he came from the West, but are not agreed as to whether he came before or after the Itzás, or with them. They say that he was well disposed, that he had not wife or children, and that after his return he was regarded in Mexico as one of their gods, and called Cezalcohuati (Quetzalcóatl). In Yucatan

---

[49] Brinton, 1882, pp. 167, 170; Roys, 1933, pp. 50, 51, 141.

[50] Ralph Roys (1933, p. 141 note 4) disagrees with this reading, translating the Maya here, *tumen Kak-u-pacal yetel Tec Uilue*, as "at the hands of Kak-u-pacal and Tec Uilu," who are known from another source (*Relaciones de Yucatan*, vol. 1, pp. 119, 269) to have been Itzá captains, said to have founded Mayapan. *Kak-u-pacal* literally means "fiery glance."

> also he was reverenced as a god, because of his great services to the state, as appeared in the order which he established in Yucatan after the death of the chiefs, to settle the discord caused in the land by their deaths.
>
> This Cuculcán, having entered into an agreement with the native lords of the country, undertook the founding of another city wherein he and they might live, and to which all matters and business should be brought. To this end he chose a fine site eight leagues further inland from where Mérida now lies, and some fifteen or sixteen from the sea. They surrounded the place with a very broad wall of dry stone some eighth of a league in extent, leaving only two narrow doorways; the wall itself was low. In the middle of the enclosure they built their temples, calling the largest Cuculcán, the same as at Chichén Itzá. They built another circular temple, different from all the others in the country, and with four entrances; also many others about them, connected one with the other. Within the enclosure they built houses for the lords alone, among whom the country was divided, assigning villages to each according to the antiquity of their lineage and their personal qualifications. Cuculcán did not call the city after himself, as was done by the Ah-Itzaes at Chichén Itzá (which means the 'Well of the Ah-Itzaes'), but called it Mayapán, meaning the 'Standard of the Mayas,' the language of the country being known as Maya. The Indians of today call it Ich-pa, meaning 'Within the Fortifications.'
>
> Cuculcán lived for some years in this city with the chiefs, and then leaving them in full peace and amity returned by the same road to Mexico. On the way he stopped at Champotón, and there in memorial of himself and his departure he erected in the sea, at a good stone's throw from the shore, a fine edifice similar to those at Chichén Itzá. Thus did Cuculcán leave a perpetual memory in Yucatan.[51]

This middle sixteenth century Spanish account of the establishment of the Itzá at Chichen Itzá, coming from the west, and the foundation of Mayapan by a great lord named Kukulcan who also came from the west, agrees remarkably well, not only with the entries in the *u kahlay katunob* for this period, but also with the archæological evidence.

The Itzá are said to have come from Chakanputun to Chichen Itzá, where they established themselves. Whether they immediately preceded, followed, or came with Kukulcan is not clear—the writer's conjecture is that the Itzá leader may well have been Kukulcan himself; the important point is that peoples coming from the southwest were beginning to enter northern Yucatan.

Landa says Kukulcan founded Mayapan while he was at Chichen Itzá, which we have seen from the *u kahlay katunob* was occupied by the Itzá in a Katun 4 Ahau, 968–987 (table 1), and since the confederacy known as the League of Mayapan began, according to the same native chronicles, in the following katun, Katun 2 Ahau, 987–1007, it would seem that Mayapan must have been founded during the last half of the tenth century.[52]

---

[51] Gates, 1937, pp. 9–11, translated from Landa, 1928–1929, pp. 60, 62, 64, 66, 68.

Finally, the *u kahlay katunob* from the Books of Chilam Balam of Mani and Tizimin both agree that Uxmal was founded by a leader named Ah Zuitok Tutul Xiu in the same Katun 2 Ahau as that in which the League of Mayapan was organized. Landa gives the same information in the following passage, further indicating that the Tutul Xiu came from Chiapas, i.e. from the southwest:

> The Indians relate that there came into Yucatan from the south many tribes with their chiefs, and it seems they came from Chiapas, although this the Indians do not know; but the author so conjectures from the many words and verbal constructions that are the same in Chiapas and in Yucatan, and from the extensive indications of sites that have been abandoned. They say that these tribes wandered forty years through the wilderness of Yucatan, having in that time no water except from the rains; that at the end of that time they reached the Sierra that lies about opposite the city of Mayapán, ten leagues distant. Here they began to settle and erect many fine edifices in many places; that the inhabitants of Mayapán held most friendly relations with them, and were pleased that they worked the land as if they were native to it.[53]

Landa's allusion here to the forty years of wandering is in perfect agreement with the forty years of wandering between Chakanputun and Chichen Itzá mentioned by the *u kahlay katunob* in both the Mani and Tizimin texts.

We see at the end of the tenth century the play of new forces in the northern Yucatan scene. Chichen Itzá, an older center, passes into the hands of a new dynasty, the Itzá, which had come from Chakanputun on the southwest coast of the peninsula. Mayapan, destined to become the political center of northern Yucatan in the second half of the Mexican Period, is founded from Chichen Itzá by a prince probably of Itzá, i.e. Mexican, origin. Finally Uxmal is founded about the same time by a leader whom the *u kahlay katunob* call Ah Zuitok Tutul Xiu, whose appelative Xiu is almost certainly of Mexican origin,[54] and who, another Spanish sixteenth century source, the *encomendero* Juan Bote, definitely states, was a native of Mexico.[55]

---

[52] Ralph Roys (1933, p. 141 note 4) concludes, probably correctly, that "the fall of Chakanputun, the establishment of Mayapan and the Itzá conquest of the important cities of northern Yucatan all took place within a single generation."

[53] Gates, 1937, p. 14, translation from Landa, 1928–1929, pp. 76, 78.

[54] Brinton, 1882, pp. 109–110.

Who then were these newcomers that took over the administration of one, and founded the two other leading city-states of northern Yucatan at the end of the tenth century, forming a confederacy known as the League of Mayapan?

Evidence, both documentary and archæological, is not wanting to indicate that at least the leaders, nobles, and priests may well have been, perhaps originally were, of Mexican origin, though if so, they and their ancestors had lived in the peninsula long enough—from two to two and a half centuries in the region of Chakanputun—to have acquired the Maya language before they occupied Chichen Itzá and founded Mayapan and Uxmal.

All the Spanish authorities agree that but one language, Maya, was spoken throughout the peninsula, though the strong Mexican flavor of the culture brought by these later groups from the southwest, as well as the specific statements that their leaders were of Mexican origin, makes it probable that originally their ancestors came from central Mexico. Even though this should be true, however, it is important to note that they had lived in the southwestern part of the peninsula long enough to have become completely Mayanized in speech before taking over the political direction of northern Yucatan.

So far as Chichen Itzá and Mayapan are concerned, probably the two first centers in the north to be established by the Maya-Mexican newcomers, the archæological evidence of a strong Mexican influence is overwhelming. This has been repeatedly reported at Chichen Itzá,[56] the greater part of which dates from the Mexican rather than the Maya Period,[57] and it has also been found at Mayapan, though to a much lesser extent.

In examining this third or Mexican Period of Yucatan history, it seems safe to conclude that during the last half of the tenth century, several groups of closely related peoples, probably of Mexican origin so far as their leadership was concerned, though all spoke Maya, entered the peninsula from the southwest and proceeded to assume the political direction of northern Yucatan, establishing Maya-Mexican dynasties at the capitals of the three leading city-states: the Itzá at Chichen Itzá, the Cocom at Mayapan, and the Xiu or Tutul Xiu at Uxmal.

---

[55] *Relaciones de Yucatan*, vol. 11, p. 287.

[56] Morley, 1913; *idem*, 1925; Morris, Charlot, and Morris, 1931; Ruppert, 1935; Pollock, 1936.

[57] Carnegie Inst. Wash., 1934, p. 91.

Perhaps as intruders, feeling the need of mutual protection in their new homes, the leaders of these three cities formed a confederacy called the League of Mayapan, under which they ruled the country jointly.[58]

There seems to have followed a period of general prosperity under the league, since two of its members, Chichen Itzá and Uxmal, grew to be the greatest cities of the New Empire, the only Class 1 sites in the northern half of the peninsula according to the writer's classification of the Maya cities.[59]

Architecture reached new levels of achievement at both sites. At Chichen Itzá the imposing pyramid-temples with their barbaric feathered-serpent columns in honor of Kukulcan, the Feathered Serpent, patron deity of the city,[60] the vast colonnaded halls,[61] and the Caracol[62] or astronomical observatory, a high round tower, are especially typical of the Mexican Period. Uxmal, on the other hand, became as typically Maya in its architecture. Indeed the Maya renaissance reached its most magnificent manifestation at Uxmal, in the Palace of the Governors—probably the finest edifice ever constructed in ancient America—and in the scarcely less meritorious Monjas Quadrangle. The Xiu capital was the greatest Maya center of the New Empire as Chichen Itzá was its greatest Maya-Mexican center.

Other sites, the very names of which are unknown, multiplied on every side, and under the peaceful conditions which prevailed during the eleventh and twelfth centuries, northern Yucatan enjoyed a prosperity literally never equaled before or since.

Suddenly and without previous warning so far as the *u kahlay katunob*, our only sources for the event, are concerned, in the tenth tun of Katun 8 Ahau, 1194 (table 1) the League of Mayapan was disrupted by a war between Chichen Itzá and Mayapan. Although the immediate cause of this conflict remains obscure, all are agreed as to its result. Hunnac Ceel, then ruler of Mayapan, with the aid of seven of his leading men, Ah Zinteyut Chan, Tzuntecum, Taxcal, Pantemit, Xuchueuet, Ytzcuat, and Kakaltecat, the names of the last four of whom, it is to be noted, are of definitely Mexican origin, defeated Chac Xib Chac, the ruler of Chichen Itzá, and drove him from the city.

---

[58] Brinton, 1882, pp. 97, 103, 141, 147; Martínez Hernández, 1927, pp. 8, 17.

[59] Morley, 1937–1938, vol. IV, p. 249, table 102.

[60] Morris, Charlot, and Morris, 1931; Carnegie Inst. Wash., 1937.

[61] Morris, Charlot, and Morris, 1931, vol. 1, pp. 52–91; Carnegie Inst. Wash., 1926, pp. 261–262; *idem*, 1927, pp. 240–241; *idem*, 1928, pp. 293–297.

[62] Ruppert, 1935.

Although the writers of the *u kahlay katunob*, describing this event, are unusually vague,[63] it seems fairly clear that Ulil, the ruler of Izamal, was also drawn into the war, perhaps on the side of Mayapan,[64] and for some years afterward there would seem to have been Itzá reprisals.[65] Perhaps the war was sectional, the East against the West.

The principal results were the disruption of the league and the centralization of supreme authority at Mayapan in the hands of the Cocom, with which may be said to have ended the first half of the Mexican Period (table 1).

Unfortunately the *u kahlay katunob* are largely silent as to the events of the next two and a half centuries, though Bishop Landa carries on the main outline of the Maya story with a little additional background material.

In order to insure the complete subjugation of the other Maya chieftains, the Cocom rulers obliged them to live within the walls of Mayapan and to administer the affairs of their respective provinces, towns, and villages through deputies, their persons in effect guaranteeing their own good behavior.[66]

Bishop Landa tells the story of this period as follows:

> The governing Cocom began to covet riches, and to that end negotiated with the garrison kept by the kings of Mexico in Tabasco and Xicalango, that he would put the city into their charge. In this way he introduced the Mexicans into Mayapán, oppressed the poor, and made slaves of many. The chiefs would have slain him but for fear of the Mexicans. The lord of the Tutul-xiu never gave his consent to this. . . . In this situation several years passed.
>
> This Cocom was the first who made slaves; but out of this evil came the use of

---

[63] Brinton, 1882, pp. 97, 102, 140–141, 147, 155, 160–161; Martínez Hernández, 1927, pp. 8–9, 16–17, 25; Roys, 1933, pp. 48, 137.

[64] Brinton, 1882, pp. 97, 102–103, 140, 146–147; Martínez Hernández, 1927, pp. 9, 17, 25–26; Roys, 1933, p. 178.

[65] Brinton, 1882, pp. 97, 103, 155, 160–161; Martínez Hernández, 1927, pp. 17, 26.

[66] Gates, 1937, pp. 15–16, translated from Landa, 1928–1929, pp. 70, 72.

> arms to defend themselves, that they might not all become slaves. Among the successors of the Cocom dynasty was another one, very haughty and an imitator of Cocom, who made another alliance with the Tabascans, placing more Mexicans within the city, and began to act the tyrant and to enslave the common people. The chiefs then attached themselves to the party of Tutul-xiu, a man patriotic like his ancestors, and they plotted to kill Cocom. This they did, killing at the same time all of his sons save one who was absent; they sacked his dwelling and possessed themselves of all his property, his stores of cacao and other fruits, saying that thus they repaid themselves what had been stolen from them.[67]

The undubitable political supremacy of Mayapan in northern Yucatan from the beginning of the thirteenth century to the middle of the fifteenth, as compared with the apparent corresponding inadequacy of its archæological remains, has prompted the Carnegie Institution during the current year to make a preliminary study of this site. This survey has just been concluded and its results are here announced for the first time.[68]

In a passage from Landa, previously quoted (page 549), it is stated that Mayapan was surrounded by a low wall of dry laid stone, an eighth of a league in extent, having two narrow doorways, that within the enclosure the temples were built, the largest being called Cuculcán, which was also the ancient name of the largest temple at Chichen Itzá, now known as the Castillo, and finally that there was a circular temple with four entrances, unlike all others in the country.[69]

The Institution's survey, however, reveals that Mayapan was a far larger site and of greater architectural importance than Landa's brief description indicates, more in keeping with the political predominance universally ascribed to it by the early authorities, both Maya and Spanish.

The wall instead of being one-third of a mile in circuit (an eighth of a league) is 5½ miles in extent, enclosing an area of about 2⅛ square miles, roughly oval in shape. It is low, only 6 or 7 feet high on the outside, including a 2-foot parapet running around the outer edge. It is made of dry laid stone and varies from 10 to 12 feet in thickness at the base. In place of the two narrow doorways mentioned by Landa, nine entrances were found, varying in width from 3½ to 10

[67] Gates, 1937, pp. 15–16, translated from Landa, 1928–1929, pp. 80, 82.

[68] This survey was made under the writer's supervision by Mr. Ralph T. Patton. A preliminary report of the results will be found in Carnegie Inst. Wash., 1938.

[69] Gates, 1937, p. 11, translated from Landa, 1928–1929, p. 66.

feet. The number nine was of unusual significance to the ancient Maya, being the number of gods who presided over the nine under worlds of Maya mythology. This, at least, raises the possibility that each of the city's nine gates may have been dedicated to one of these under-world deities. Several stairways ascend the wall on the inside. Finally, instead of the single round tower described by Landa, the Institution's expedition located four of these circular constructions within the walled area.

The city was abundantly supplied with water, no less than nineteen *cenotes* being found inside the walls. The heaviest concentration of buildings is near the center of the walled area, though minor constructions occur throughout the enclosure.

Landa also mentions seven or eight stones

> well carved and bearing lines of the characters (hieroglyphics) they use . . . which are thought to be a monument of the foundation and destruction of the city . . . the natives being asked what they were, answered that it was the custom to set up one of these stones every twenty years [every katun] that being the number by which they reckon their periods.[70]

These stones were stelæ, of which nine sculptured ones have been found at Mayapan. It has been possible to decipher the Maya dates of two of these probably, and of a third doubtfully. Stela 1 would seem to have been erected to commemorate the end of a Katun 10 Ahau, probably 10.18.0.0.0 *10 Ahau* 3 Tzec (?), 1185; Stela 6, probably the end of a Katun 13 Ahau, 11.3.0.0.0 *13 Ahau* 18 Pax (?), 1283; and Stela 5, less surely, the end of a Katun 4 Ahau, 11.1.0.0.0 *4 Ahau* 8 Mol (??), 1244; all within a century of each other, and perhaps dating from the last half of the Mexican Period (table 1).

Landa in a passage already quoted (pages 553–554) describes forcefully the misrule of the Cocom during the thirteenth and fourteenth centuries, and tells how in order to force their will upon the other lords of Yucatan they brought from Xicalango in Tabasco Mexican soldiers who were permanently garrisoned at Mayapan. Thus they were able to quell the rising discontent, enslave many, and oppress all.

The Tutul Xiu do not seem to have been involved in the war against Chichen Itzá at the end of the twelfth century (page 553), which probably explains why the Cocom, backed by their Mexican allies, were able to usurp the political leadership of Yucatan from that time forward—to the victors went the spoils. Landa states clearly that the Tutul Xiu never gave their consent to these tyrannies, however, and further that the head of the family in the middle of the fifteenth century was "a patriotic man like his ancestors."

---

[70] Gates, 1937, pp. 16–17, translated from Landa, 1928–1929, p. 84.

Finally in 1441, goaded beyond further endurance by the Cocom tyranny, the Maya chieftains, uniting under the leadership of the Tutul Xiu, attacked Mayapan, sacked the city, and slew its ruler and all his sons except one who was absent on a trading expedition to Honduras at the time. The victors each withdrew to their respective provinces, towns, and villages, and centralized government in Yucatan came to an end, closing the second half of the Mexican Period (table 1).

Bishop Landa, by means of an elaborate calculation in which he lists the various calamities which afflicted northern Yucatan during the next century and a quarter, i.e. that immediately preceding the writing of his *Historia de las Cosas de Yucatan* in 1566,[71] fixes the date of the fall and subsequent abandonment of this city as having taken place in 1441.

Taking the year of his writing, 1566, as his point of departure for working backward, he says:

> Since this last plague more than fifty years have now [1566] passed, the mortality of the wars was twenty years prior, the pestilence of the swelling was sixteen years before the wars, and the hurricane other sixteen years before that, and twenty-two or twenty-three years after the destruction of the city of Mayapán. Thus according to this count it has been 125 years since its overthrow within which the people of this country have passed through the calamities described.[72]

This fixes the date of the plague as in 1516 or possibly 1515, the date of the mortality due to the wars as 1496, the date of the pestilence as 1480, the date of the hurricane as 1464, and the date of the destruction of Mayapan as 1441. The *u kahlay katunob* from the Book of Chilam Balam of Tizimin and the first and second chronicles from the Book of Chilam Balam of Chumayel record that there was a pestilence in Katun 4 Ahau, 1480–1500 (table 1), i.e. at the very beginning of this katun according to Bishop Landa. Again, the first *u kahlay katunob* from the Book of Chilam Balam of Chumayel describes an epidemic of smallpox as having taken place in Katun 2 Ahau, 1500–1520. This is doubtless Landa's pestilence "with great pustules that rotted the body,"[73] which occurred according to his reckoning in 1515 or 1516. Small though these two items of confirmation may be, they indicate nevertheless the high degree of reliability of the native chronicles, our principal documentary sources for the reconstruction of Maya history.

---

[71] The only copy of Landa's invaluable history that has survived is a manuscript in the former Royal Academy of History at Madrid, which bears the date 1566.

[72] Gates, 1937, p. 19, translated from Landa, 1928–1929, pp. 94, 96.

All but one of the five *u kahlay katunob* (the third from the Book of Chilam Balam of Chumayel) and two other early Spanish sources, Cogolludo writing in 1656[74] and Villagutierre Soto-Mayor in 1700,[75] are in agreement that the fall of Mayapan took place in a Katun 8 Ahau, 1441–1461 (table 1), and since Landa's calculations indicate very clearly that this event occurred in 1441, the sources may all be said to fix the destruction of this city as having taken place at the very beginning of this Katun 8 Ahau.

But Landa in another passage gives us a third chronological fact of high confirmatory value. He says:

> . . . after having been established in this city [Mayapan] for more than five hundred years, they abandoned and left it desolate, each going to his own country.[76]

According to the *u kahlay katunob*, Mayapan was founded between Katuns 6 Ahau and 2 Ahau, 948–987 (table 1), and further since five centuries prior to 1441 takes us back to about 941 it will be seen that Landa and the native chronicles are also in substantial agreement as to the date of the foundation of this city.

The breakdown of centralized authority in northern Yucatan was speedily followed by corresponding political disorganization. As a direct result of the fall of Mayapan, all the larger cities were abandoned.

The Itzá left Chichen Itzá, retiring far to the south to Lake Petén Itzá in what is now the Department of Petén, Guatemala, establishing their new capital, Ta Itzá or Tayasal, in the same Katun 8 Ahau, on a promontory at the western end of the lake.[77] The Chels, a prominent noble family of Mayapan, after the fall of the city, established their principal settlement at Izamal. The single surviving son of the slain Cocom ruler gathered the remnants of his people about him and established his rule at Tibolon[78] near Sotuta. Even the victorious Tutul Xiu did not return to Uxmal but founded a new capital which they called Mani, meaning in Maya "it is finished."[79]

---

[73] Gates, 1937, p. 19, translated from Landa, 1928–1929, p. 94.

[74] Cogolludo, 1688, p. 507.

[75] Villagutierre Soto-Mayor, 1701, pp. 29, 106, 514.

[76] Gates, 1937, p. 16, translated from Landa, 1928–1929, p. 82. Herrera (1726–1730, vol. 4, p. 207) gives the same figure: "i desempararon la Ciudad, deseando cada Señor vivir en libertad en sus Pueblos, al cabo de quinientos Años, que se fundò, en la qual havian vivido con mucha Policia."

[77] Villagutierre Soto-Mayor, 1701, p. 31.

Yucatan fell apart into a score of petty provinces, which, urged on by ancient feuds and jealousies, waged incessant war against each other. These conflicts, added to the divers calamities described by Landa, reduced the country to political chaos and paved the way for its final conquest by the Spanish in 1541. Weakened as they were, however, the Maya managed to withstand the attacks of their better-armed invaders for fourteen years before they finally succumbed.

The defeat of a coalition of Maya chieftains in 1541 and the foundation of the Spanish capital at the site of Merida on January 6, 1542 (Julian) brought about the final subjugation of the northern half of the peninsula.

The Itzá, who had withdrawn to the south a century earlier and had established themselves at the western end of Lake Petén Itzá, managed to survive for another century and a half, until finally conquered by the governor and captain general of the Provinces of Yucatan, the Sergeant Major Martín de Ursúa y Arizmendi, in 1697, an event which brought to an end the last independent Maya political entity.

It is yet too early to permit final conclusions as to whether general laws underlay and controlled the development of Maya culture from its humble beginnings during the first millennium B.C. to its final disintegration in the middle of the fifteenth century, or whether, like the familiar Topsy of American literature, Maya culture just "growed," influenced, coerced, and conditioned by environment, it is true, but subject to the operation of no more general principles of historical development.

An interesting recurrent pattern, however, a confirmation from ancient American sources of the familiar adage that history repeats itself, is to be found in the thirteen centuries of the Maya story reviewed here.

The first period of Maya cultural florescence, the Maya Old

---

[78] Gates, 1937, p. 17, translated from Landa, 1928–1929, pp. 84, 86, 88.

[79] Herrera, 1726–1730, vol. 4, p. 208.

Empire, may be said to have dated in terms of their own chronology from the beginning of Baktun 9, A.D. 435. The period of Old Empire culture roughly coincides with this baktun, which, practically considered, witnessed its rise, apogee, and decline.

Baktun 9 began, or more strictly speaking the last katun of the preceding baktun, Baktun 8, ended, on the day 8 Ahau. In much later New Empire times the priestly compilers of the *u kahlay katunob* selected katuns which ended on this same day as the points at which the sequence of the katuns was considered to have begun to repeat itself. This is clearly described in the Codice Perez,[80] where these Katuns 8 Ahau are called the *uuɔ katun*, "the doubling or folding over of the katuns," i.e. the points at which the thirteen differently named katuns began to repeat themselves, which took place every 13×20=260 tuns or about 256¼ years.

The writer believes that this particular katun was so chosen as the point at which the katun sequence was supposed to have begun, because the first great outburst of Maya culture took place in a baktun (no. 9) that was counted from a baktun (no. 8) which ended on the day 8 Ahau.

Owing to the accidental fact that the baktun 9.0.0.0.0, with which the *u kahlay katunob* begins, happens to have ended on this day, as time went by the ceremonial importance of Katuns 8 Ahau became so great that it finally influenced, if indeed it did not largely determine, the general pattern of later Maya history.

Referring to table 1 it will be seen that the first recurrence of a katun ending on the day 8 Ahau after the beginning of the sequence was in 672–692, when Chichen Itzá is said to have been heard of for the first time. This, however, could hardly have been more than a coincidence, certainly not due to any psychological connection between time and event.

The second return of Katun 8 Ahau fell in 928–948, when the *u kahlay katunob* are in all but unanimous agreement that Chakanputun was abandoned and the Itzá migrated northeastward. Here may well have been the beginning of chronologic coercion. Katun 8 Ahau had come back again. The katun sequence was about to renew itself. The times were ripe for change. The Itzá had already been at Chakanputun for more than two centuries. Under the Maya system of agriculture the land must already have begun to yield decreasing returns. What could have been a more propitious time to set out upon a search for a new home?

---

[80] Brinton, 1882, pp. 84–85, translated from the Codice Perez, pp. 90, 168.

The third return of Katun 8 Ahau, 1185–1204, is marked by an event of clearly indicated deliberation, the plot of Hunnac Ceel, the ruler of Mayapan, and the war against Chichen Itzá. The writer suggests this particular katun may well have been chosen for this important venture since by this time the Maya records clearly set forth that Katuns 8 Ahau were times of change, perhaps even peculiarly auspicious therefor. Chichen Itzá is attacked, its ruler defeated, and Mayapan emerges from the war with the hegemony of northern Yucatan in her hands, a resounding justification of the potency of the psychologic moment.

The fourth return of Katun 8 Ahau, 1441–1461, marks another epochal event in Maya history, again one shaped by deliberate design: the plot of the Maya chieftains in 1441 under the leadership of the Tutul Xiu against the supreme ruler of northern Yucatan, the Cocom at Mayapan; the attack on that city; the death of its ruler and the subsequent political disintegration of the country. Here was every opportunity for free advance selection of the most auspicious date upon which to "break" the revolt. Katun 8 Ahau had returned. As their own history had repeatedly demonstrated, the portent of the period was one of change. Only now the exigency of the times ran against the plotters; Cocom tyranny could no longer be endured. Scarcely had the favorable katun begun, i.e. in its very first year, when desperate necessity precipitated the event, long overdue, which had been deferred only till the moment was ceremonially favorable. Their chronological beliefs indicated change was due and straightway the Maya made history—their last collective action before the Spanish conquest.

Coming finally to the post-Conquest period, by one of those curious turns in the wheel of national destiny, the last act of the Maya drama fell just short of the fifth return of Katun 8 Ahau, 1697–1717.

Cogolludo reports that when Fathers Bartolomé de Fuensalida and Juan de Orbita visited Tayasal, the last independent Maya capital, in the middle of October 1618, and attempted to evangelize the Itzá, Canek, the Itzá ruler, told them that the time had not yet arrived when their ancient priests had prophesied that they were to abandon the adoration of their native gods:

> To this the Canek replied that the time had not yet arrived in which their ancient priests had prophesied that they should leave off the adoration of their gods, because the age [katun] in which they were at present [October 1618] was that which they call Ox Ahau (which is to say third age [Katun 3 Ahau]) and that there had not yet arrived what was appointed for them.[81]

Villagutierre Soto-Mayor is more explicit, since by inference at least he seems to indicate that this abandonment of their old pagan cult was to take place in a Katun 8 Ahau:

> To this Canek replied: That although there had not arrived this time in which their ancient priests had prophesied that they were to leave off the adoration of their gods; because the age in which at this time they were was that which they call Oxahau, which is to say Third Age [Katun 3 Ahau]: (these barbarians seem to have counted their Ages backward, or to a determined number; which reached they forgot it, and returned to begin the count anew [the *uuɔ* katun]: Because when they retired from Yucatan, which already is going on for three hundred years,[82] they say that it was the Eighth Age [Katun 8 Ahau] and that there had not arrived so soon what was appointed for them; and now [1618] they say it was the Third Age and that the time had not yet arrived.)[83]

The writer suggests that the inference here is that the "appointed time" was Katun 8 Ahau.

In the middle of January 1696, Father Andrés de Avendaño y Loyola visited Tayasal, in another attempt to evangelize the Itzá and persuade them to accept the authority of the king of Spain. Avendaño, as he himself states, was well acquainted with the Maya calendar and reminded the Itzá how according to their own prophecies the time was approaching when it had been predicted they were to give up the worship of their pagan deities and accept Christianity. The Itzá, while again admitting these prophecies, still contended that the time ordained for this to be done had not yet arrived. Finally, after further conferences, it was agreed that four months were still wanting before the appointed period, though the best correlation of Maya and Christian chronology indicates that at the time of Avendaño's visit (January 1696) fully a year and a half was still lacking before the beginning of the next Katun 8 Ahau.

[81] Cogolludo, 1688, pp. 494–495. Reference to table 1 will show that in the middle of October 1618 Katun 3 Ahau had just begun, Katun 5 Ahau, which also marked the end of baktun 12.0.0.0.0, having closed on September 20, 1618.

[82] Villagutierre Soto-Mayor is correct in his time estimate here, since he wrote in 1699 or 1700 and the fall of Mayapan and the withdrawal of the Itzá from Chichen Itzá took place in 1441, i.e. more than two and a half centuries before.

[83] Villagutierre Soto-Mayor, 1701, pp. 105–106.

But again time was forcing Maya destiny. In March of the following year (1697), Martín de Ursúa y Arizmendi in command of a force of Spaniards and Indians from Yucatan finally pushed through to the shores of Lake Petén Itzá. A galley having been built to transport the army to Tayasal at the western end of the lake, the final attack was made on March 13. In spite of overwhelming numbers, perhaps as high as five thousand Indians against one hundred and eight Spaniards, a single discharge of musketry was sufficient to dispel the Itzá hosts, which fled in every direction abandoning their belongings, their homes, even the sanctuaries of their tribal deities, almost without resistance.

But why? What the reason for this precipitous flight practically without striking a blow once the battle had been joined? On the day of the attack there lacked only 136 days before the beginning of Katun 8 Ahau, at which time, the writer believes, their ancient prophecies had long foretold their end would come. The Itzá, fearfully convinced of their approaching doom, not only predicted by their native prophets but also indicated by their meticulously kept historical records, gave up the unequal struggle against their own hostile gods and fled, thus anticipating by less than five months the last return of this most fateful katun of Maya history.

LITERATURE CITED

Bartlett, H. H., Paul C. Standley, E. B. Mains, Joyce Hedrick, W. R. Taylor, C. V. Morton, J. R. Swallen, E. C. Leonard, and E. P. Killip, 1935. Botany of the Maya Area: Miscellaneous Papers I–XIII. Carnegie Inst. Wash. Pub. No. 461.

Benedict, F. G., and Morris Steggerda, 1936. The Food of the Present-Day Maya Indians of Yucatan. *Contr. Amer. Archæol.*, vol. 3, no. 18, Carnegie Inst. Wash. Pub. No. 456, pp. 155–188.

Brinton, D. G., 1882. The Maya Chronicles, *Library of Aboriginal American Literature*, no. 1. Philadelphia.

Carnegie Institution of Washington, 1916. Carnegie Inst. Wash. Year Book No. 15.

——— 1920. Carnegie Inst. Wash. Year Book No. 19.

——— 1924. Carnegie Inst. Wash. Year Book No. 23.

——— 1925. Carnegie Inst. Wash. Year Book No. 24.

——— 1926. Carnegie Inst. Wash. Year Book No. 25.

——— 1927. Carnegie Inst. Wash. Year Book No. 26.

——— 1928. Carnegie Inst. Wash. Year Book No. 27.

——— 1930. Carnegie Inst. Wash. Year Book No. 29.

——— 1932. Carnegie Inst. Wash. Year Book No. 31.

Carnegie Institution of Washington, 1933. Carnegie Inst. Wash. Year Book No. 32.

——— 1934. Carnegie Inst. Wash. Year Book No. 33.

——— 1937. El Castillo, Pyramid-Temple of the Maya God, Kukulcan, *Carnegie Inst. Wash. News Serv. Bull.*, vol. 4, no. 12.

——— 1938. Carnegie Inst. Wash. Year Book No. 37.

Cogolludo, Diego Lopez, 1688. Historia de Yucathan. Madrid.

Grandz, Solomon, 1936. Review of History of Hindu Mathematics by B. Datta and A. N. Singh. *Isis*, vol. 25, pp. 478–488. Bruges.

Gann, T. W. F., 1926. Ancient Cities and Modern Tribes: Explorations and Adventure in Maya Lands. London.

——— 1935. Tzibanché, Quintana Roo, Mexico, *Maya Research*, vol. 2, no. 2, pp. 154–166.

Gates, William, 1937. Yucatan, before and after the Conquest, by Friar Diego de Landa, with other related documents, maps and illustrations, translated with notes by William Gates. Maya Society, Pub. No. 20, Baltimore.

Gaumer, George F., 1937. Monografía de los Mamíferos de Yucatán. Mexico City.

Herrera, Antonio de, 1726–1730. Historia General de los Hechos de los Castellanos en las Islas y Tierra Firme del Mar Oceano. 5 vols. Madrid.

Howe, George P., 1911. The Ruins of Tuloom, *Amer. Anthropologist*, n. s., vol. 13, pp. 539–550.

Landa, Diego de, 1928–1929. Relation des choses de Yucatan (Relación de las cosas de Yucatan). Texte espagnol et traduction française en regard. Edition complète accompagnée de notes tirées des principaux auteurs anciens et modernes précédée d'une introduction sur la vie et l'oeuvre de l'auteur et suivie d'un index des matières contenues dans la Relation, par Jean Genet. 2 vols. Paris.

Lizana, Bernardo de, 1893. Historia de Yucatan. Devocionario de Ntra. Sra. de Izmal y Conquista Espiritual. Museo Nacional de Mexico, Mexico City.

Lothrop, S. K., 1924. Tulum: An Archæological Study of the East Coast of Yucatan. Carnegie Inst. Wash. Pub. No. 335.

Lundell, C. L., 1937. The Vegetation of Petén. Carnegie Inst. Wash. Pub. No. 478.

Martínez Hernández, Juan, 1927. Crónicas Mayas. Merida.

Mercer, Henry C., 1896. The Hill Caves of Yucatan: A Search for Evidence of Man's Antiquity in the Caverns of Central America, Being an Account of the Corwith Expedition of the Department of Archæology and Palæontology of the University of Pennsylvania.

Morley, Sylvanus G., 1913. Archæological Research at the Ruins of Chichen Itzá, Yucatan. *Reports upon the Present Condition and Future Needs of Anthropology* presented by W. H. R. Rivers, A. E. Jenks, and S. G. Morley, Carnegie Inst. Wash. Pub. No. 200, pp. 61–91.

——— 1915. An Introduction to the Study of the Maya Hieroglyphs. U. S. Bur. Amer. Ethnol. Bull. No. 57, Smithsonian Inst.

Morley, Sylvanus G., 1917. The Ruins of Tuloom, Yucatan: The Record of a Visit of the Carnegie Institution Central American Expedition 1916 to an Important but Little Known Ancient Maya City, *Amer. Mus. Jour.*, vol. 17, no. 3, pp. 190–204.

——— 1917*a*. Rise and Fall of the Maya Civilization in the Light of the Monuments and the Native Chronicles, *Proc. Internat. Cong. Americanists, XIX Session, Washington 1915*, pp. 140–149.

——— 1920. The Inscriptions at Copan. Carnegie Inst. Wash. Pub. No. 219.

——— 1925. Chichen Itzá, An Ancient American Mecca: Recent explorations in Yucatan are bringing to light the temples, palaces and pyramids of America's most holy native city, *Nat. Geog. Mag.*, vol. 47, no. 1, pp. 63–95.

——— 1927. New Light on the Discovery of Yucatan and the Foundation of the New Maya Empire, *Amer. Jour. Archæol.*, 2d ser., vol. 31, no. 1, pp. 51–69.

——— 1935. Guide Book to the Ruins of Quirigua. Carnegie Inst. Wash. Supp. Pub. No. 16.

——— 1937–1938. The Inscriptions of Petén. Carnegie Inst. Wash. Pub. No. 437. 5 vols.

Morris, Earl H., Jean Charlot, and Ann Axtell Morris, 1931. The Temple of the Warriors at Chichen Itzá, Yucatan. Carnegie Inst. Wash. Pub. No. 406, 2 vols.

Murie, Adolph, 1935. Mammals from Guatemala and British Honduras, Mus. Zool., Univ. Michigan, Misc. Pub. No. 26, pp. 7–30.

Pearse, A. S., and collaborators, 1938. Fauna of the Caves of Yucatan. Carnegie Inst. Wash. Pub. No. 491.

——— E. P. Creaser, and F. G. Hall, 1936. The Cenotes of Yucatan: A Zoological and Hydrographic Survey. Carnegie Inst. Wash. Pub. No. 457.

Pollock, Harry E. D., 1936. The Casa Redonda at Chichen Itzá, Yucatan. *Contr. Amer. Archæol.*, vol. 3, no. 17, Carnegie Inst. Wash., Pub. No. 456, pp. 129–154.

Redfield, Robert, and Alfonso Villa R., 1934. Chan Kom, a Maya Village. Carnegie Inst. Wash. Pub. No. 448.

Relaciones de Yucatan, 1898–1900. Relaciones Histórico-Geográficas de las Provincias de Yucatán, *Colección de Documentos Inéditos Relativos al Descubrimiento, Conquista y Organización de las Antiguas Posesiones Españolas de Ultramar*, ser. 2, vols. 11, 13.

Roys, Ralph L., 1931. The Ethno-Botany of the Maya. *Middle Amer. Research Ser.*, Pub. No. 2, Tulane Univ. of Louisiana.

——— 1933. The Book of Chilam Balam of Chumayel. Carnegie Inst. Wash. Pub. No. 438.

Ruppert, Karl, 1935. The Caracol at Chichen Itzá, Yucatan, Mexico. Carnegie Inst. Wash. Pub. No. 454.

Shattuck, George C., 1933. The Peninsula of Yucatan: Medical, Biological. Meteorological and Sociological Studies. Carnegie Inst. Wash. Pub. No. 431.

Standley, Paul C., 1930. Flora of Yucatan. *Bot. Ser.*, vol. 3, no. 3, Field Mus. Nat. Hist. Pub. No. 279.

STEGGERDA, MORRIS, 1932. Anthropometry of Adult Maya Indians: A Study of Their Physical and Physiological Characteristics. Carnegie Inst. Wash. Pub. No. 434.

TEEPLE, JOHN E., 1930. Maya Astronomy. *Contr. Amer. Archæol.*, vol. 1, no. 2, Carnegie Inst. Wash. Pub. No. 403, pp. 29–115.

THOMPSON, J. ERIC, 1929. Maya Chronology: Glyph G of the Lunar Series, *Amer. Anthropologist*, n. s., vol. 21, No. 2, pp. 223–231.

——— 1931. Archæological Investigations in the Southern Cayo District, British Honduras. *Anthropol. Ser.*, vol. 17, no. 3, Field Mus. Nat. Hist. Pub. No. 301.

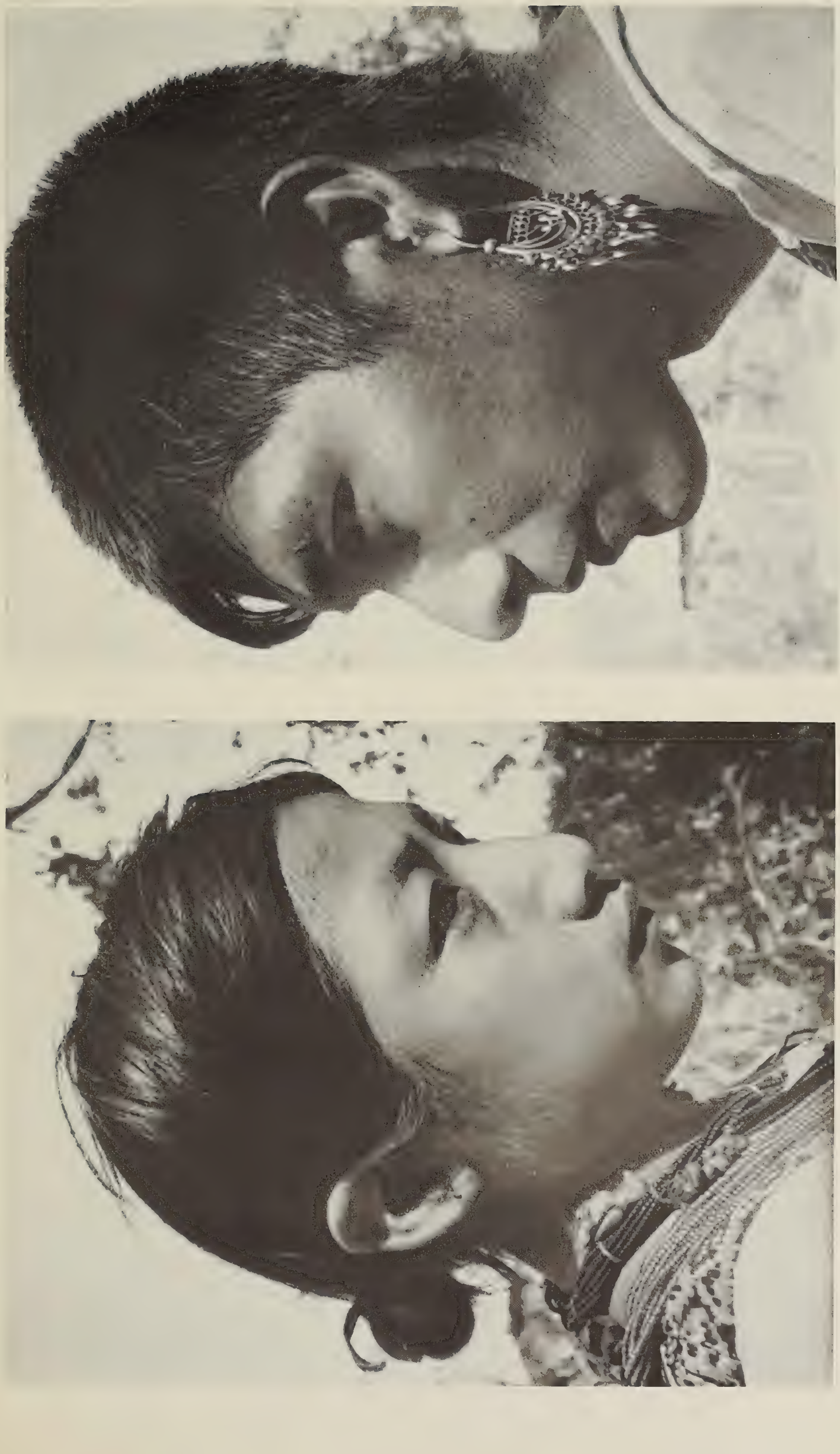
——— HARRY E. D. POLLOCK, and JEAN CHARLOT, 1932. A Preliminary Study of the Ruins of Coba, Quintana Roo, Mexico. Carnegie Inst. Wash. Pub. No. 424.

VAN TYNE, JOSSELYN, 1935. The Birds of Northern Guatemala. Mus. Zool., Univ. Michigan, Misc. Pub. No. 27.

VILLA R., ALFONSO, 1934. The Yaxuna-Cobá Causeway. *Contr. Amer. Archæol.*, vol. 2, no. 9, Carnegie Inst. Wash. Pub. No. 436, pp. 187–208.

VILLAGUTIERRE SOTO-MAYOR, JUAN DE, 1701. Historia de la Conquista de la Provincia de el Itzá, Reduccion, y Progressos de la de el Lacandon, y Otras Naciones de Indios Barbaros de la Mediacion de el Reyno de Guatimala a las Provincias de Yucatan, en la America Septentrional. Primera Parte. Madrid.

WILLIAMS, GEORGE D., 1931. Maya-Spanish Crosses in Yucatan, *Papers Peabody Mus. Amer. Archæol. and Ethnol.*, Harvard Univ., vol. 13, no. 1, pp. 1–256.



Maya from southeastern Yucatan

THE MAYA INDIANS OF YUCATAN

MORRIS STEGGERDA

Department of Genetics
The Carnegie Institution of Washington

APPROXIMATELY twenty-five kilometers south of Chichen Itza, Yucatan, is a settlement of five houses called Xmax. In those houses live twenty or twenty-five Indians, who are separated from other settlements by miles of dense, thorny, dry, tropical jungle. When the author passed through Xmax in the late afternoon of a hot day, the men, wearing only cotton trunks, were seated outside their thatched houses, obviously enjoying pleasant conversation. The women and children stood still and appeared uninterested as two American travelers with their guides and pack horses disrupted momentarily their monotonous, tranquil existence. In the short time that it took our little pack train to walk fifty meters, we had passed through a settlement and were again on our way in the tall bush. As an image is caught by a photographic shutter, that scene at Xmax was imprinted on the mind of the writer. It typifies graphically the life and existence of more than two-thirds of the Indian population of Yucatan.

These little clearings in the thick, high bush, the thatched pole houses built up around the all-important *cenote* or water hole, and the small, outlying corn patches (*milpas*), where trees have been felled exposing the rocky soil; these would dot the map of Yucatan if all such settlements could be shown. Thus, the *cenote* and its immediate surroundings will supply the few necessities which this little group of people may need for five years or perhaps fifty or five hundred years. The changing wet and dry seasons each demand their specific work. During the dry season from November to March, the bush is cleared and burned to make a *milpa*. With the coming of the rains in May, corn is planted, and once or twice during the growing season, the fast-growing weeds are cut. In September the green corn is eaten, and in October and November enough maize is harvested from the nine- or ten-acre field to supply the average family of five with enough food for the remainder of the year, with a surplus sufficient to buy gunpowder for the men and cotton cloth and colored thread for the women. These, together with salt and *machetes* (knives for cutting the bush), represent practically their only imports.

Xmax is a community in the heart of the Yucatan jungle that is making a success of life. In it people are eating, working, resting, and reproducing under conditions which for most white men would be almost intolerable. The rocks, the shallow soil, the snakes, the wild animals, the harmful insects, the thorns and briars of Xmax have been conquered by a contented and happy family group of Maya Indians, and Xmax is but one of the hundreds of tiny settlements that dot the Yucatan bush. The government census, published in 1934,[1] listed 1655 inhabited communities, of which 66 per cent had fewer than 50 inhabitants; 18 per cent had populations ranging between 50 and 200; 11 per cent were between 200 and 1000; 4 per cent between 1000 and 10,000. There were only two towns, or 0.1 per cent, listed as having over 10,000 individuals. Thus small towns like Xmax are characteristic of the Yucatan bush today.

This ratio of small to large towns may well duplicate the conditions found in ancient times. Perhaps then the Indians, living in temporary thatched houses, made pilgrimages to the markets and shrines of Chichen Itza, Uxmal, Coba, and Izamal, and other towns, as they do today to such centers as Merida, Valladolid, and Tizimin.

The author has spent the equivalent of two years with the present-day Maya, working and visiting in about fifty of their towns and villages. He has measured and tested more than a thousand individuals. From notes gathered, and from averages of measurements taken on these Maya from the interior, it is possible now to answer questions concerning their size, teeth, metabolism, and food. We can make statements concerning the average number of children in a family, telling how many are born, how many reach maturity, the average age at death, and the chief causes of death. We can describe their homes, farms, and animals, their personalities, and their way of life, and we can present a few considerations concerning the present-day Maya in relation to their famous ancestors. With this outline in mind, let us introduce to you the typical Maya Indian of Yucatan.

---

[1] Quinto Censo de Poblacion, Mexico, D. F., 1934. For complete reference see "Literature Cited" at end of paper.

The present-day Maya Indian of Yucatan is the result of wide racial mixture. If we assume that an indigenous population existed in Yucatan before A.D. 500, we may also assume that it was influenced biologically as well as culturally by the immigrants from the so-called "Old Empire," since this is universally the case when two human stocks meet on a common ground. These indigenous people of northern Yucatan undoubtedly were also influenced biologically by related tribes from southern Mexico.[2] After the first hundred years of Spanish occupation, the genetic stock of the Yucatan Maya must have shown considerable change, for not only was white Spanish blood introduced into their strain, but also small quantities of Negro blood, for at one time during the early colonial period there were 134 Negro slaves in Merida alone. Today only rarely does one see a full- or even a half-blooded Negro in Yucatan, but one does occasionally find Indians with curly hair, which may have been derived from Negro ancestors. In colonial times, settlements were established by Whites other than Spanish, e.g., Irish and French. In recent times, Chinese and Korean men have been allowed to enter Yucatan, but their wives have been barred; thus a strain of Oriental blood has been introduced. Today, while there are localities where the people seem relatively free of race mixture, it is in the larger centers of population that such mixture is most evident. Our studies were, therefore, limited to the small settlements, largely in the eastern part of the peninsula where the population was most nearly pure Indian.

The present Yucatan Maya are very short people. They are the "pygmies" of the American Indians. Among Negroes there are the short Bushmen and the tall Bantus, and among Whites there are the short Mediterraneans and the tall Scandinavians. Just so, among the American Indians, one finds the short Maya of southern Mexico and Central America, and the tall Indians from the Great Plains. The average stature for adult Yucatan Maya males is 155 cm, or 5 feet, 1 inch.[3] They are shorter than other Indians when they are born and remain shorter throughout their years of growth as well as when they are adults. Their growth pattern, however, is similar to that of other races. In the white race, the rate at which the average individual grows decreases during the years from approximately six to eleven years. After this period an acceleration in the rate of growth occurs until the age of 14.5 years. At this age, although growth continues, the rate at which it occurs, decreases. This same growth pattern occurs among Negroes and Navajos and also among these Maya. Thus for each of these races, the rate of increase in male stature is lowest at 10.5 years and highest at 14.5 years. This is extremely interesting, since members of these races were all measured by the author with a similar technique, and were from entirely different habitats. The Whites were from a Dutch community in Michigan, the Negroes from Alabama, the Navajos from Arizona, living at an altitude of at least 5000 feet, while the Maya live in the subtropics at sea level.

---

[2] Morley, 1938.

[3] M. Steggerda, 1932.

Maya children have relatively large chests compared with Navajos, Whites, and Negroes. Children of each of these races have a lower relative chest girth at the ages of ten to fourteen than they have at six or eighteen years of age. That is, children just previous to the adolescent spurt of growth are all relatively slender. The Maya children are least slender; at thirteen years of age, the average chest girth for Maya boys is 51.4 per cent of their stature. For Navajo boys the ratio of these dimensions is 48.3 per cent. Dutch Whites, who are still more slender, have a chest girth which at thirteen years is 47.9 per cent of their stature, while Negroes, who are most slender, have a relative chest girth of 46.3 per cent. These differences are all statistically significant. For adult males these percentages of the relative chest girths are greater. For example, the average chest girth of an adult male Jamaica Negro is 49.9 per cent of his stature;[4] in Dutch Whites,[5] it is 53.4 per cent; and in the Maya, 56.0 per cent. These figures show that the Maya, both children and adults, are very short and thickset.

The arms of the Maya are long in relation to their stature, and the shoulder widths of the adults equal those of their much taller fellow Indians from the Great Plains. The Maya have broad heads; they are classified as brachycephalic, having a cephalic index of 85 per cent. Their noses are generally broad, pronounced, and curved. Their faces likewise are broad. Their hair is black and straight, their eyes uniformly dark brown or black. In many individuals, the eye is characteristically Mongolian; that is, the epicanthic fold is pronounced. The Maya Indians greatly resemble the Chinese in such features as face, hair, and eyes, and also in less conspicuous characteristics, such as palmar dermatoglyphics or the ridges on the palm.[6]

---

[4] Davenport and M. Steggerda, 1929.

[5] M. Steggerda, 1932*a*.

At the base of each finger of every palm there is an area where three of these ridge lines meet. That meeting place at the base of each finger is called a tri-radius, and the proximal line which runs from this tri-radius is traced to the periphery of the palm. The periphery of the palm has been arbitrarily numbered from 1 to 13. The line which runs from the tri-radius at the base of the fifth finger is called line D. From the tri-radius of the fourth finger, line C extends to some part of the periphery of the palm. Likewise, line B extends from the middle finger, and line A from the tri-radius at the base of the index finger. Among white people, the main line formulation generally conforms to an 11.9.7.5 type, which means that line D terminates in position 11, line C in position 9, line B in 7, and line A in 5. For Negroes, this same main line formulation generally reads 7.5.5.5, and for Chinese, the formulation is 9.7.5.5. The Maya follow, in the main, the 9.7.5.5 type of formulation for these main lines. At least, the 9.7.5.5 type is more characteristic than the other types of formulæ.

The Maya are also like the Chinese in other dermatoglyphic characteristics; for example, the unusually vertical configuration of the palm as indicated by the proximal termination of line A. This means that for Chinese and the Maya, line A terminates quite frequently in position 1 or 2, while among Whites and Negroes, this line terminates more often in position 5, giving the palm a more transverse or oblique configuration. Further, line C is often absent or greatly reduced in both Chinese and Maya. This is not so characteristic among the Whites and Negroes. These similarities between Chinese and Maya are true for two other palm ridge characteristics, all of which indicate a remarkable degree of racial similarity between these two groups. These studies on the palms of the Maya were carried on with the cooperation of Inez D. Steggerda, the author's wife. Closely associated work on the finger[7] and the sole[8] prints, collected by the author, was done by Dr. H. Cummins of The Tulane University.

---

[6] I. D. Steggerda, M. Steggerda, and Lane, 1936.

Investigations of the physical features of the Yucatan Indian led to a detailed study of the teeth of the Maya.[9] Generally speaking, the Maya's teeth are exceptionally good. Our study considered the condition of the deciduous or milk teeth and also of the adult or permanent teeth. For the deciduous teeth of the children, our study included 707 Maya between the ages of two and fourteen years. For comparison, the author examined the mouths of 1082 Navajo children from Arizona and 970 Dutch White children in Michigan. In the two Indian groups, only 50 per cent of the children are afflicted with caries or tooth decay in their deciduous teeth, as contrasted with 90 per cent in the Dutch group. Not only do more Dutch children have caries, but the number of cavities per mouth for all ages is greater. The percentages are, for the Dutch, 4.8; Navajo, 2.8; Maya, 2.3. Data on the permanent teeth pertain to even larger numbers and include a fourth racial group, the Negroes. For this adult group, 944 Maya mouths were examined, 1358 Dutch, 1861 Navajos, and 599 Jamaica Negroes, or a total of 4762 individuals.

From this study it was learned that the number of people afflicted and the percentage of teeth involved increase with advancing age. In both Maya and Navajo races, more than 50 per cent of the individuals are free from dental caries until the age of twenty; in the Dutch White group more than half have dental caries in their permanent teeth by the age of nine, and more than 90 per cent by the age of fourteen.

In the studies of caries on Whites, a certain bacterium, *Bacillus acidophilus*, has been found to be particularly prevalent in mouths where much decay is present. In order to ascertain the relative presence of this bacterium in the mouths of the Maya, who live on a high carbohydrate diet and who at the same time have a low incidence of caries, a bacteriological examination of their saliva was made. Dr. T. J. Hill, professor of oral pathology at Western

---

[7] Cummins and M. Steggerda, 1936.

[8] Cummins and M. Steggerda, 1936*a*.

[9] M. Steggerda and Hill, 1936.

Reserve University, made the analysis, working with the author at Chichen Itza and Piste. He concludes his study with these words:

> Of the one hundred eighty people examined, 42 or 23.3% had no cavities and had lost no teeth. Of the group examined, 34 people had previously been examined by Morris Steggerda in 1931. In the six years which had elapsed since the previous examination, ten of these had developed no caries, fourteen had developed one cavity each and the remainder had an average of 2.2 cavities per person. Of the one hundred and eighty persons examined, 48 persons or 26.6% had negative cultures from both quantitative and qualitative examination for B. acidophilus. The numerical count varied in the mouths of different persons from zero to 85,000 per cubic centimeter of saliva. The average for the group which at some time had active caries was 5,900, while the average for the group which had never had caries was 400 B. acidophilus per cubic centimeter of saliva. The average B. acidophilus count for the 34 people examined in 1931 and having developed an average of 2.2 cavities per person in six years was 500 per cubic centimeter.
>
> The Maya Indians are relatively free from dental caries, compared with the White race. The incidence of people affected and the percentage of teeth involved is nearly identical with that found in the Navajo Indian. This is of unusual interest because the high carbohydrate diet of the Maya and the equally high protein diet of the Navajos may have a relationship to the difference in incidence of caries upon the smooth surfaces of the teeth of these races. While there is much variation in the calcification of teeth, microscopic sections of several Maya teeth do not show them to be better calcified than the average teeth from the White race.
>
> One might expect that the high carbohydrate diet would be conducive to the growth of B. acidophilus, but these organisms, while found to be present, are not found as frequently in the mouths of the Maya Indians as in the White race, nor is their numerical presence as great. The relationship of the presence of B. acidophilus to the presence of caries in both the individual Maya, and this race as compared to the White race, suggest that, to a certain extent, resistance to dental caries may be inherited as an individual or racial characteristic.

As we have said, the diet of the Maya is chiefly carbohydrate.[10] This statement was based on a three-year study of the food of the Maya, made in collaboration with Dr. F. G. Benedict, former director of the Nutrition Laboratory of the Institution. The most common food items found in the diet of the Maya are maize, beans, other native vegetables, and a little meat, but more than 75 per cent of their food is corn. The method of its preparation as well as the secondary items of the Maya diet are practically the same today as they were in Landa's[11] time, 1566, when he wrote:

---

[10] Benedict and M. Steggerda, 1936.

> Their principal sustenance is maize, of which they prepare various dishes and drinks; and even drunk as they do it, it serves as their food and drink. The Indian women put the maize to soak the night before in lime and water, and in the morning it is soft and half-cooked, having in the process lost the husk and nib. They next grind it on stones, and when half ground make it into great balls and loads for the use of laborers, travelers and sailors. In that shape it keeps for several months, except for souring. Of this they then take a lump and dissolve it in a vessel or gourd formed by the rind of a fruit that grows on a tree, whereby God has provided them with vessels; out of this they drink the liquor and then eat the rest, it being of excellent taste and very nourishing. (*atole*) From the maize that is more fully ground they take away the milk and thicken it at the fire, making a sort of curd for morning use; and this they drink hot. (*pozole*) Upon what is left from the morning they put water for drinking through the day, since they are not accustomed to drink water alone.
>
> Out of maize and ground cacao they make a sort of froth that is very delicious, and with which they celebrate their festivals. From the cacao they extract a grease that is much like butter, and from this together with maize they prepare another agreeable and much esteemed drink. Still another is made from the substance of the ground maize raw, which is also fresh and tasty. They prepare many kinds of bread, good and healthful, except that it is not good to eat when cold; so that the Indian women are kept busy with making it twice a day. (*tortillas*) They have not learned how to make flour that can be kneaded like wheat flower, and when they do make it as one makes wheat bread, it is good for nothing. They make ragouts of vegetables and venison, and of wild and domestic fowls of which there are plenty, and of fish of which there are plenty. Thus they have good provisions, especially since they are raising Spanish pigs and fowls.
>
> The men and women do not eat together, but eat apart, on the ground; or if there is much, with a mat for a table. They live well when they have it, and endure hunger equally well when they have not, getting along on very little. After eating they wash their hands and mouth.

Though this account was written 350 years ago, it still applies in the main today. The contemporary Yucatan Maya boil their corn in lime water, drink *atole* and *pozole*, and make their *tortillas* twice a day. The men and women still eat apart, and after eating, wash their hands and rinse out their mouth.

Concerning the food of the present Maya it can be said that on the average, 75 per cent of their daily energy intake comes from maize. The daily protein intake per individual averaged 74 grams, a rather low figure. Analyses of 24-hour collections of urine of four male Maya for nitrogen content indicated that the nitrogen excretion per kilogram of body weight was not significantly different from the average found in a similar study on male medical students in the United States. The total daily energy intake averaged 2565 calories, which is low when compared with the daily energy intake (3500 calories) of the average American laborer.

---

[11] Friar Diego de Landa, 1549–1579. See Gates, 1937, p. 34.

The scientific study of the quality and quantity of food eaten by the Maya was also undertaken because two Harvard expeditions[12] in collaboration with the Carnegie Institution of Washington had previously put forth some evidence that the Maya Indians of Yucatan had a higher metabolic rate than the norms obtained for white people in the United States. Other races previously studied had lower rates than Whites. These included Negroes,[13] Chinese,[14] Australians,[15] and Indians from India.[16] During the field season of 1931–1932, the author took metabolic readings on 30 adult Maya men, each of whom was tested three times a day for three successive days, making a total of 90 individual experiments. The results [17] were analyzed by Dr. F. G. Benedict of the Nutrition Laboratory of the Carnegie Institution. The total heat production, or metabolism, of these thirty men when compared with the Harris-Benedict predictions for white men of the same age, weight, and height, was found in the majority of cases to be higher than the predicted rate, making the percentage deviations plus. Of these percentage deviations, there were only four minus values, and of these, two were −4.0 per cent, and the other two were less than −1.0 per cent. Of those subjects having a higher metabolism than the prediction, one showed a deviation of +26.1 per cent, and 11 of the other 29 men gave values of over +10 per cent. The average for all 30 was +8.4 per cent.

Because of the uniformity of the diet of the Maya, the author is frequently asked about their diseases, and whether these Maya are, as a whole, healthy or sickly. If there is any positive statement to be made about the Maya, it is that they are not sickly, although there is a very high infant mortality. The adults are strong and robust. This strength in adulthood may be the result of the survival of only the strong individuals. However, the Maya do have sicknesses, and when compared with highly civilized communities, they do die early. Concerning some vital statistics of the Yucatan Maya, the author can supply the following:

---

[12] Williams and Benedict, 1928; Shattuck and Benedict, 1931.

[13] M. Steggerda and Benedict, 1928.

[14] Benedict and Meyer, 1933.

[15] Hicks, Matters and Mitchell, 1931.

[16] Mason and Benedict, 1931.

[17] M. Steggerda and Benedict, 1932.

In 1936, the four Yucatan towns, Piste, Pencuyut, Chan Kom, and Xocenpich all together had 265 family units, or households. This generally included a father, mother, children, and occasionally an aunt, grandmother, or nephew. In these 265 households, the average number of inhabitants was 4.95, or practically five persons in each. Forty-seven Maya mothers produced, on the average, 8.14 children each. The age at the first menstruation was 12.85 years for 125 Yucatecan girls (most of whom were Indian). This figure is of interest since it represents cases for which we were able to secure a birth date. Thus, we can say with some assurance that the first menses occurs in Yucatan before thirteen years of age. We were able also to secure birth dates and marriage dates for 36 Maya girls, for whom the average age at marriage was 16.3 years. The average age at the birth of the first child for 57 married women for whom we secured birth dates was 17.8 years. This compares exactly with material supplied by Aberle[18] on Pueblo women in the United States. The average length of reproductive life of 39 women in the two towns of Piste and Pencuyut was found to be approximately 17 years.

The death rates for Piste over the same years average 31.29 per thousand; Pencuyut is slightly higher, with a rate of 33.78 per thousand; and Dzitas is considerably lower than the other two towns, having an average rate of 23.39 per thousand. These death rates are inaccurate, since the death records for a given village often include those from surrounding hamlets, whereas such people are not usually considered in the town's population.

The birth and death rates for Piste are approximately three times as high as those for the United States. In 1876, when the birth rate in Europe was at its maximum, Germany led European countries with a rate of 40.9 per thousand. This was more than 20 points less than the average birth rate in Piste.

The death rate in Piste is also high as compared with European countries. Again, according to figures for 1876, the highest death rate was in Italy, and was 28.8.[19] The death rate in the Maya village of Chan Kom, as reported by Redfield,[20] was 26.4 per thousand. For Merida, Valladolid, and Ticul, as reported by Shattuck,[21] the death rates for the years 1924 to 1928 were respectively 35, 30, and 29 per thousand.

---

[18] Aberle, 1931, p. 76.

In Piste the total number of deaths at one year of age or less, not including stillbirths, for the period from 1918 to 1935, was 82. In this period there were 368 births, giving a rate of 223 infant deaths per thousand births. For Pencuyut from 1902 to 1935 there were 121 infant deaths in 785 births, or a rate of 154 per thousand. For Pueblo Indian children of Arizona and New Mexico in the United States, as reported by Aberle,[22] the rate is 228.5 per thousand. She reports from the literature a rate of 154.4 per thousand for United States colored children and 108.3 for United States white children. In Piste, the sex ratio for infant mortality is 57 per cent for males and 43 per cent for females; in Pencuyut, the ratio is 48 per cent for males and 52 per cent for females.

Up to the present time any data published on the mortality of the Yucatan Indians are subject to error because of the inaccuracy in reporting the age of the deceased. To determine the error in the Yucatan death records we attempted to secure the exact birth dates of at least 50 individuals of each age group for whom we also know an exact death date. Thus, by the comparison of ages as given in the death records and the actual ages as determined from both birth and death dates, it was possible to compute the error of the ages as given in the death records. The data were gathered chiefly from the records of Piste, Pencuyut, and Dzitas. Fifty-two per cent of the ages given for the group from five to nine years were correct, while the remainder deviated from the correct age from −6 to +2 years. For another example, in the age group from thirty to thirty-four years, only 16 per cent of the estimates of the age at death were correct; the other 84 per cent ranging in error from −12 to +6 years. In the thirty-five- to thirty-nine-year group, the error ranged from −19 to +23 years, while in the forty-five to forty-nine age group, the error ranged from −15 to +9 years. This shows the trend of the errors between the real and given age at the time of death. Thus in computing the life span of an individual it is necessary to consider the average error for each age group.

---

[19] The figures for the birth and death rates in Germany and Italy are from the *Encyclopedia Britannica* (14th ed., 1929), vols. 3 and 7.

[20] Redfield and Villa, 1934, p. 13.

[21] Shattuck, 1933, p. 387.

[22] Aberle, 1932, p. 344.

In Yucatan the chief cause of death is described as "fever." This, of course, is a vague term, and may be ascribed to a great number of specific diseases. Diarrhea and dysentery come next, followed by pellagra, pneumonia, and influenza. The data were derived from the stated causes of death in the Piste, Pencuyut, and Dzitas records. A few records were copied in the towns of Cuncunul and Chapab as well. Whooping cough, malaria, and measles cause 6 per cent, 4 per cent, and 3 per cent respectively of the total number of deaths. Accidents are in the same frequency as measles. We secured the stated causes of death for 1057 individuals. Measles ends with the twenty-five-year class and is highest in the ten- to fourteen-year group. Pellagra, on the other hand, increases in percentage with age. Fever as a cause of death is highest among the younger groups, relatively low during middle life, but increasing with old age. This is also true for pneumonia and influenza. Diarrhea and dysentery remain at approximately the same percentage throughout the life span.

Whether the Maya family is rich or poor, the house is very much the same. Wauchope[23] in writing of Maya houses says, "I believe that the types of 'bush house' seen almost everywhere throughout the Maya area today have changed very little in the last 1900 years."

The present-day "bush-house" (one having a thatched roof) is rectangular in form, with rounded ends, and has an average length of 25 feet and a width varying from 10 to 12 feet. The walls, about 6 feet high, are made of poles from 1 to 2 inches in diameter. Sometimes the spaces between the poles are daubed with mud. Four large upright poles support the palm-thatched roof. There are two doors, one in each long wall. In the construction of the house there is no need of nails since lianas (woody vines) are used to fasten parts together. The floors are generally bare earth, sprinkled with water daily to settle the dust.

On the floor at one end of the house, generally away from the prevailing easterly wind, there are three stones marking the boundary of the family fire. A stone *metate* for grinding the corn is generally set on a long, low bench near by. Many families use a small

[23] Wauchope, 1934, p. 123.

metal hand mill for grinding the corn, but recently the stone grinder and the hand-operated mills are being replaced in some of the larger towns by the gasoline-operated mills run by private individuals or occasionally by the town as a community project.

In the Maya household, nearly everything is made by hand, usually by the head of the house. This includes the small tables, chairs, washtubs, gourd dishes, pottery dishes, and water jugs. Small benches or hand-hewn blocks of wood serve as chairs. A three-legged table upon which *tortillas* are made is generally found in each house. The henequen hammocks and ropes, bridles, and native saddles are also handmade.

At the end of the room opposite the fireplace stands the family trunk, which holds extra clothing and other possessions. Next to it, corn may be stored, either loose on the floor, or in sacks or bins. Chickens, pigs, and dogs have free run of the house; naturally fleas and ticks are common.

There are no pictures on the walls, no clocks, no mirrors; however, in many homes one finds a table which serves as an altar. On it generally stands a cross with an image or a picture of a saint, with candles and cloth draperies which lend an air of sanctity. In the few Protestant homes in Piste, the open Bible takes the place of the cross and candles.

Hung along the walls or stuck between the poles of every dwelling are a shotgun, an ax, a *machete*, one or more water gourds, and a small henequen bag called a *zabucan*. A crude shelf along the wall holds salt, spices, and whatever china dishes the family may own.

Hanging from the cross-poles which support the walls and roof are a number of hammocks. These are the beds of Yucatan, and the number in each household depends somewhat upon the number of members, since often several persons may sleep in one hammock.

As protection from insects and rodents, a number of objects are hung from the ceiling. Most commonly suspended in this fashion are bundles of beans, still in the pod. These serve as food and also as seed for the coming year. Dishes made from half gourds, containing unused food, are hung from the ceiling by cords which pass through an inverted gourd, preventing rats and mice from running down the strings to get at the food in the dish.

The Maya Indians of Yucatan farm in a unique manner. Because of the shallowness of the soil, averaging three or four inches in depth, they cannot hoe, plow, or cultivate as do the corn farmers of the United States. They have a system known as *milpa* farming, in which each farmer cuts the trees and shrubs from an area which he selects. After these cut trees have dried for several months, the field is burned. This burning occurs in March or April just previous to the rainy season. After the rains begin, generally in May, the corn is planted, and the weeds and corn grow up together. Once or twice during the growing season the weeds and sprouting tree stumps are cut back with the *machete*. When the kernels have set on the ear and the corn has matured, the stalks are bent over so that the ears of maize hang downward. This is done to prevent the birds from eating the kernels and to keep the grain dry during succeeding rains, since in this hot, humid climate it would otherwise become moldy. The soil, although shallow, is very fertile, and during the three or four summer months, the corn grows to a height of 12 to 15 feet, producing on an average the equivalent of 16 to 25 bushels per acre. With all expenses paid, and at current prices for labor, the *milpero*, or corn farmer, in Yucatan averages approximately a peso and a half profit for each *mecate* (20 m square), which is the equivalent of $4.22 per acre. The average *milpa* for the Chichen Itza area is 99.4 *mecates*, or approximately 9 acres.

In the small Indian towns and hamlets, the men themselves generally carry the corn to and from their *milpas*; some, however, own one or two horses. In Piste, in 1934, a town having 364 people, there were 61 horses and 82 cows. The average family keeps one or two pigs and a flock of about fifteen chickens.

If one enters the plaza of a Yucatan Indian town at noontime, or even in the cool of an evening, the quietness and seeming peacefulness of the village may cause him to desire to spend a few years of his life there in relaxation, saying to himself, "Nothing happens here, everyone is my friend, let me stay here and join this society and rest." This idea of complete rest and tranquillity soon disappears as the traveler becomes better and more intimately acquainted with the people. For soon he will witness the effects of jealousy among the young married couples, and the bitterness engendered by old family feuds, and he will experience the tenseness of every *milpero* as he scans the sky in anticipation of too much or too little rain. He will see the anxiety in nearly every home over an ever-present sick member, and he will observe the hope of cure through some mystic powers of a *yerbatero*. The heat of the day and the coldness of the night, the fleas, ticks, lice, bedbugs, mosquitos, wasps, flies, and the rats, snakes, and the wild animals of the bush may not be great factors in influencing his mental attitude while in this Yucatan society, but they will cause discomfort and some anxiety. He will talk to the people about them, and soon will combat them in their own native way. Nearly all Maya are subject to an enormous number of superstitions, fears, and taboos. For example, they believe it will bring them bad luck if they see a tarantula, a crab, or a scorpion carrying its young, and nearly everyone can cite examples to prove the validity of these contentions. Evil spirits, bad winds, evil eyes, and black magic are powerful influences in the minds of the Yucatan Maya. The traveler will become acquainted with most of them, and as he becomes still better acquainted with each family, he will learn of quarrels and of old-time grudges and even murders, which are avenged sometimes long after the occurrence of the crime. Thus, the traveler who becomes friends with these people soon finds himself with some of the same mental attitudes as the inhabitants, but chiefly those of anxiety and fear. At present, the Maya do not fear neighboring tribes, for they are not a warlike people, but they do fear various supernatural forces. The dominating influence in their life is agricultural. Therefore they are greatly dependent upon nature, which brings their chief happiness as well as their greatest fears.

The description of mental traits presented in this paper is the result of the author's experience with the Maya, and is quantitatively corroborated by the opinion of thirty people of his own background. These thirty people, all of whom were well acquainted with the Yucatan Maya, were asked to form an opinion concerning various mental traits. These opinions were assembled, and where similarity was reached, the author assumed that there was some significance to the opinion. From this approach to the subject of the mental characteristics of the Maya, we can say that when compared with American Whites, they are clean, thrifty, and conservative. In contrast with Navajo Indians, they are quick, active, alert, polite, and expressive. They have a practical sense of humor, their sex instinct is not overemphasized, but they do have a tendency to excessive alcoholism. It can be said also that this tendency toward excessive alcoholism cannot be attributed to their associations with white civilization. Toward strangers, the Maya are kindly, and among themselves they are sociable.

Some questions concerning the mental characteristics of the Maya were also answered by the *mestizo* and Spanish classes of Yucatan, as well as by the Maya themselves. From a study of all the ratings it would seem that the white raters had had a tendency toward grading the Maya too high, i.e., producing a "halo effect." For example, 65 per cent of the white raters think of the Maya as being unusually clean. This seems to be true also for all writers describing the Maya from Landa's[24] time to the present. The *mestizos* grade the Maya only as fairly clean, and Maya graders likewise classify their fellow townsmen as only fairly clean.

The consensus of the white people grading the Maya is that they are humorously inclined, whereas the *mestizos* and Maya believe that they are only moderately humorous. Concerning sociability, the whites believe them to be socially inclined; the *mestizos* consider them as rather unsociable, while the Maya classify themselves as moderately sociable.

However, these differences are only slight, and the descriptions set forth in this article, as well as in the more detailed report on the mental characteristics of the Maya, are not meant to be what the *mestizos* think of the Maya, nor what the Maya think of themselves, but rather the consensus of opinion as to what white people think of the Maya.

Discussing more of these mental characteristics of the present Yucatan Maya, it may be said that they seem very strongly attached to their children, and the children when of age respect sincerely the rights of their parents, especially those of their mothers. Abortion and infanticide are relatively unknown, and all married couples are desirous of having children, and parents rarely punish children for their misdeeds.

The Maya are not aggressive. In ancient times they were known as a peace-loving agricultural people, as is very characteristic of them today. The people are not unduly ambitious. The Maya do not desire luxuries. Their houses are all very much the same. No one demands the comfort of a bathtub or a toilet. They do not have the feeling of "keeping up with the Mex's." Maya families are not greedy for more land, more corn, more cattle, or more horses. In Piste (a typical town) no Indian has more than two horses, very few have as many as ten cows, but all have plenty to eat. A hunter shoots only enough for his family. Rarely do they steal from one another, and outwardly everyone seems contented. The Yucatan Maya are chiefly of one class, all are agriculturalists; there are no idle rich, none are ambitious business men, and none are shiftless beggars. All are hard-working men, and each earns his living from year to year.

---

[24] Landa wrote concerning the Maya Indians of Yucatan who lived in the latter part of the sixteenth century. This work has recently been published for the first time in English; see Gates, 1937.

Regarding their sex behavior, the Indians are exceedingly moderate. Incest is not common, and rape is practically unknown. Sexual intercourse before marriage occurs occasionally without undue notice; generally the boys and girls marry young, under twenty years of age, and are generally faithful to one another throughout their lives. This is of particular interest in view of the standard set by the Spanish and *mestizo* class. Among the Maya, "old maids" and bachelors are exceedingly rare.

These Indians are not given to suicide; they rarely take their own lives, and when they do, they are usually mentally deranged, by pellagra or excessive alcoholism.

The Maya existed in Yucatan long before the coming of the Spaniards, and in some parts of the peninsula are existing today relatively free from white man's influence. If all white influence could be removed from the life of the Maya, he would carry on with success as did his ancestors before him.

LITERATURE CITED

Aberle, S. B. D., 1931. Frequency of Pregnancies and Birth Interval among Pueblo Indians, *Amer. Jour. Phys. Anthropol.*, vol. 16, no. 1, pp. 63–80.

——— 1932. Child Mortality among Pueblo Indians, *Amer. Jour. Phys. Anthropol.*, vol. 16, no. 3, pp. 339–349.

Benedict, F. G., and M. H. Meyer, 1933. The Basal Metabolism of American-born Chinese Girls, *Chinese Jour. Physiol.*, vol. 7, no.1, pp. 45–49.

——— and M. Steggerda, 1936. The Food of the Present-day Maya Indians of Yucatan. *Contr. Amer. Archaeol.*, vol. 3, no. 18, Carnegie Inst. Wash. Pub. No. 456, pp. 155–188.

Censo De Poblacion (Quinto), 15 De Mayo de 1930, Estado De Yucatan. Mexico, D. F., 1934.

Cummins, H., and M. Steggerda, 1936. Finger Prints in Maya Indians, *Measures of Men*, Pub. 7, Dept. Middle Amer. Research, Tulane Univ. of Louisiana, New Orleans.

Cummins, H., and M. Steggerda, 1936*a*. Plantar Dermatoglyphics in Maya Indians, *Maya Research*, vol. 3, nos. 3–4.

Davenport, C. B., and M. Steggerda, 1929. Race Crossing in Jamaica. Carnegie Inst. Wash. Publ. No. 395.

Gates, William, 1937. Yucatan, before and after the Conquest, by Friar Diego de Landa, with other related documents, maps and illustrations, translated with notes by William Gates. Maya Society, Pub. No. 20, Baltimore.

Hicks, C. S., R. F. Matters, and M. L. Mitchell, 1931. The Standard Metabolism of Australian Aboriginals, *Australian Jour. Exper. Biol. and Med. Sci.*, vol. 8.

Mason, E. D., and F. G. Benedict, 1931. The Basal Metabolism of South Indian Women, *Indian Jour. Med. Research*, vol. 19.

Morley, S. G., 1938. The Maya New Empire, pp. 533–565 in the present volume.

Redfield, R., and A. Villa R., 1934. Chan Kom, a Maya Village. Carnegie Inst. Wash. Pub. No. 448.

Shattuck, G. C., 1933. The Peninsula of Yucatan: Medical, Biological, Meteorological and Sociological Studies. Carnegie Inst. Wash. Pub. No. 431.

——— and F. G. Benedict, 1931. Metabolism of Maya Indians in Yucatan, *Amer. Jour. Physiol.*, vol. 96, no. 3, pp. 518–528.

Steggerda, I. D., M. Steggerda, and M. S. Lane, 1936. A Racial Study of Palmar Dermatoglyphics, with Special Reference to the Maya Indians of Yucatan, *Measures of Men*, Pub. 7, Dept. Middle Amer. Research, Tulane Univ. of Louisiana, New Orleans.

Steggerda, M., 1932. Anthropometry of Adult Maya Indians—A Study of Their Physical and Physiological Characteristics. Carnegie Inst. Wash. Publ. No. 434.

——— 1932*a*. Physical Measurements on Dutch Men and Women, *Amer. Jour. Phys. Anthropol.*, vol. 16, no. 3.

——— and F. G. Benedict, 1928. The Basal Metabolism of Some Browns and Blacks in Jamaica, *Amer. Jour. Physiol.*, vol. 85, no. 3, pp. 621–631.

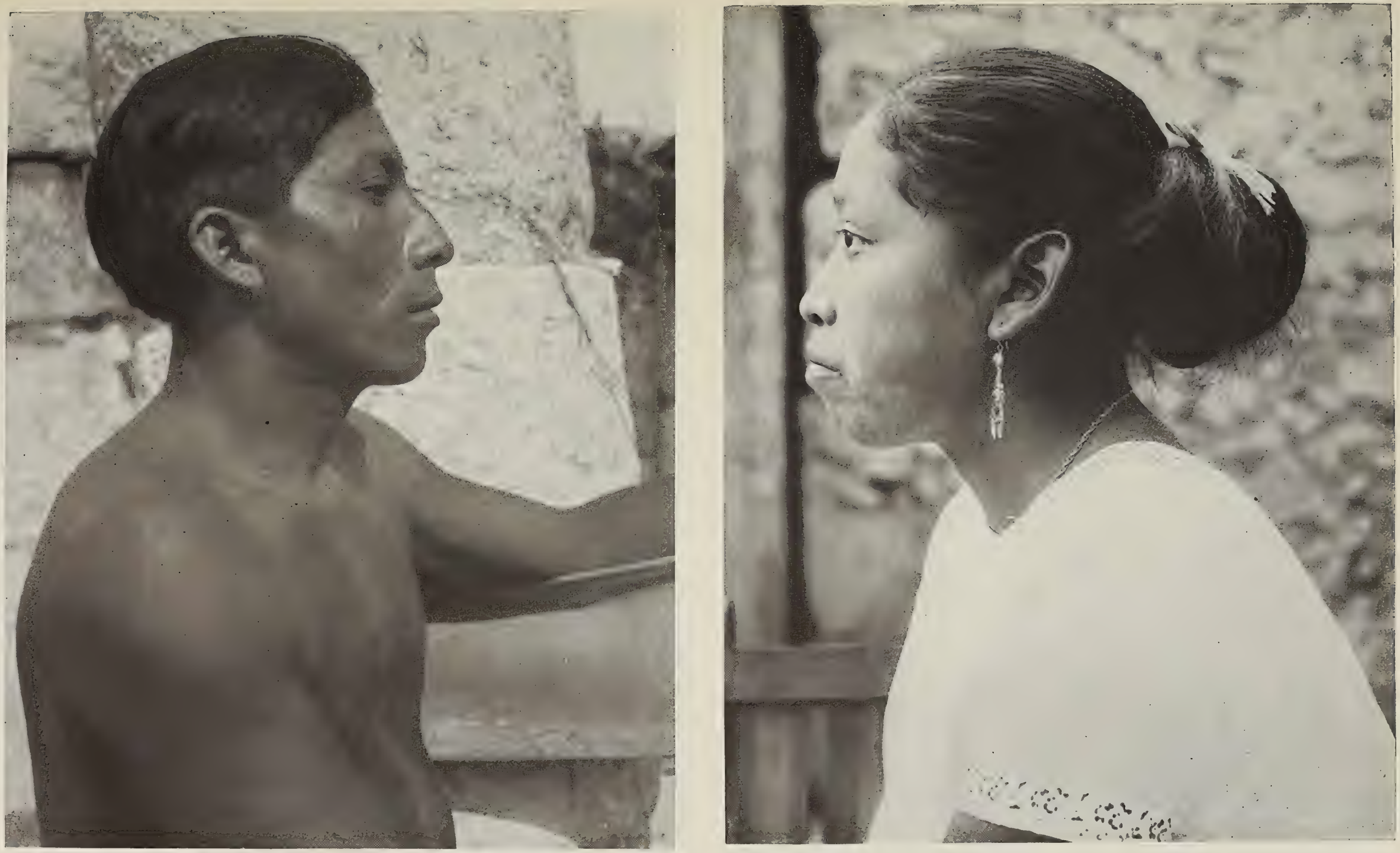
——— ——— 1932. Metabolism in Yucatan—A Study of the Maya Indian, *Amer. Jour. Physiol.*, vol. 100, no. 2, pp. 274–284.

——— and T. J. Hill, 1936. Incidence of Dental Caries among Maya and Navajo Indians, *Jour. Dental Research*, vol. 15, no. 5.

Wauchope, Robert, 1934. House Mounds of Uaxactun, Guatemala. Contr. Amer. Archæol., vol. 2, no. 7. Carnegie Inst. Wash. Pub. No. 436, p. 123.

Williams, G. D., and F. G. Benedict, 1928. The Basal Metabolism of Mayas in Yucatan, *Amer. Jour. Physiol.*, vol. 85, no. 3, pp. 634–649.

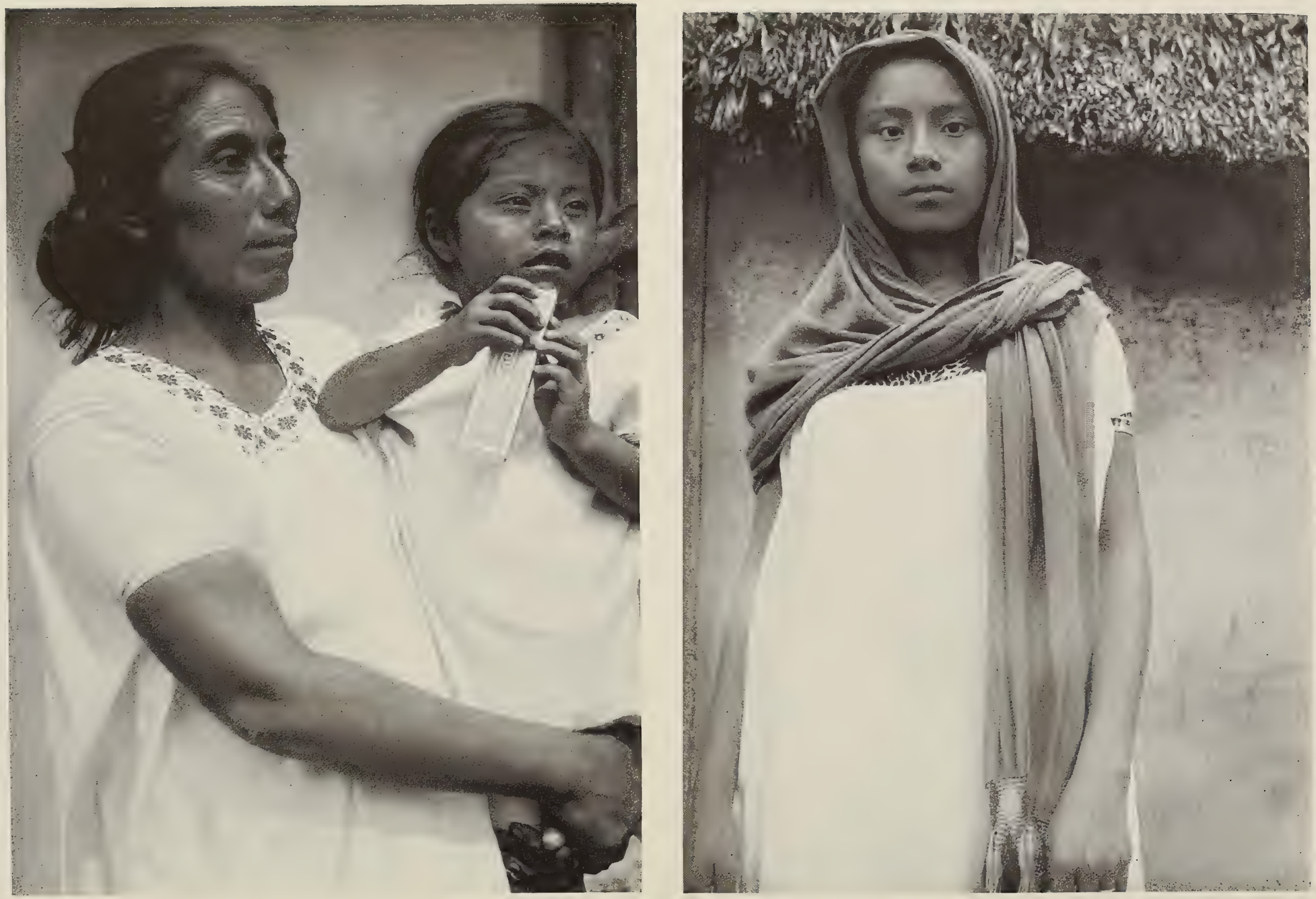
[STEGGERDA] PLATE 1



MAYA INDIANS FROM THE CHICHEN ITZA AREA

Facundo Ceme, 1932

Maxima Mex, 1934



MAYA INDIANS FROM THE CHICHEN ITZA AREA

Sabina Dzib and grandchild, 1931

Cresencia Dzib, 1937
(fourteen years of age)

FRAY DIEGO DE LANDA AND THE PROBLEM OF IDOLATRY IN YUCATAN

FRANCE V. SCHOLES AND RALPH L. ROYS
Division of Historical Research
The Carnegie Institution of Washington

THE introduction of Christianity in Yucatan subsequent to the Spanish occupation of the peninsula had serious repercussions on the traditional character of Maya culture. The age-old forms of native religious belief and ceremonial practice were closely integrated with every form of human activity. Among the deities in the Maya pantheon were gods and goddesses of rain and the winds, agriculture, commerce, and the hunt, medicine, war, and death. Maya learning consisted of a complex body of myth and legend, calendrical knowledge, and ritual, of which the priests were the chief custodians, interpreters, and teachers. Each year a series of important religious festivals were celebrated, beginning with the elaborate New Year rituals, which were followed in turn by a monthly round of lesser ceremonies related to special phases of native life, such as hunting, bee keeping, maize culture, or medicine. Human sacrifice, offerings of birds and animals, and various forms of self-mortification constituted the most striking elements of ceremonial practice, but no less important as an illustration of the powerful influence of religion was the simple rite of burning incense performed each night by Indian travelers, "praying to the god they called Ekchuah, that he might bring them safely back home."[1] Maya religion was essentially a way of life. Consequently the forced acceptance of the Christian faith produced a conflict of loyalties that had a powerful influence on the subsequent evolution of Yucatecan society.

I

The effective beginnings of the missionary program in Yucatan date from 1545, when two groups of Franciscans, one from Mexico and the other from Guatemala, arrived in the peninsula. Monastic houses were immediately founded in Campeche and Merida which became centers for the teaching and indoctrination of the Indians in the near-by pueblos. Additional friars were sent to the province from time to time during the succeeding twenty years, and the missionary area was slowly expanded. By 1563 twelve Franciscan foundations had been established and small churches had been built in a larger number of pueblos. Although a few members of the secular clergy also served in certain areas, the major burden of the missionary program was carried by the friars.

---

[1] Landa, 1937, p. 46.

During this early period slow progress was made in teaching the Indians the elements of the Christian faith. In order to facilitate the missionary program the friars, acting in conjunction with the civil authorities, concentrated the Indian settlements into larger units or removed them from outlying areas to villages near the monastic houses. Sons of caciques and other chieftains were assembled in mission schools for special instruction. At first the clergy devoted their efforts mostly to the teaching of a few simple prayers, such as the Pater Noster and the Ave Maria, to the enforcement of attendance at mass and other religious services, and to giving instruction concerning the importance, character, and implications of the sacraments, especially baptism, marriage, and extreme unction.

The clergy faced certain difficulties, however, in the successful prosecution of even this minimum program. Although some of the friars, notably Fray Luis de Villalpando and Fray Diego de Landa, learned to speak and write Maya and gave instruction to the others, it is doubtful whether more than half of the clergy became proficient in the language. Moreover, the total number of missionaries was never sufficient for the needs of the native population, which at that time numbered about three hundred thousand.[2] The province contained more than two hundred pueblos, and most of these could be visited only occasionally and for very brief periods of time. Consequently the day-to-day instruction of the Indians had to be left to native schoolmasters who had received a little instruction in the elements of Christian doctrine.

In the pueblos with permanent resident clergy, the management of local affairs was rapidly subjected to the directing influence of the Church. The missionaries intervened in the election of local officials and participated in the administration of village finances. Realizing that the success of their labors depended in part on good treatment of the Indians, they took active measures to combat abuses, such as excessive demands for labor and tribute by the colonists. Spaniards who set an evil example by immoral and irreligious conduct were subjected to ecclesiastical pains and censures. These activities inspired resentment among the Spanish ruling classes, who bitterly complained that the clergy infringed on secular interests and intervened in affairs touching the civil jurisdiction. As a result of this conflict of economic and religious motives in provincial life, a series of unseemly incidents occurred that reduced the effectiveness of the missionary program.

---

[2] This estimate is based on data in the tribute roll of 1548–1549. Archivo General de Indias, Guatemala, 128.

From the beginning the missionaries and the civil authorities waged an unceasing campaign against Maya religion and other native customs that might tend to undermine the loyalty of the Indians to the new faith. Human sacrifice was prohibited as constituting homicide, and other forms of sacrificial rites were also forbidden as heretical and contrary to Christian teaching. Efforts were made to subject the celebration of Indian festivals and dances to rigid control and to introduce Christian themes as the basis of those that were permitted. Tattooing, body painting, the use of ear- and nose-plugs, and the wearing of masks and other ceremonial costumes were forbidden. As a means of imposing Christian concepts of marriage, stringent penalties were decreed for parents who sold their daughters in marriage and for husbands who followed the ancient custom of deserting their wives because they were barren. Likewise, the clergy tried to combat the old system of employing *casamenteros*, or matchmakers, and in their teachings they stressed the degrees of carnal and spiritual relationships within which unions were forbidden by canon law. As a condition of baptism, caciques were often forced to free their slaves, many of whom had been acquired as concubines or as victims for sacrificial rites.[3]

Although these changes were in the interest of a higher form of culture and were merely practical applications of Christian moral and social concepts, they constituted a radical reorientation of native life, and the Indians were naturally reluctant to accept them. The most stubborn resistance came from the caciques, *principales*,[4] and native priests, or *ah-kins*, who realized that the new order of things threatened their traditional authority and influence as leaders of the Maya. The functions of the *ah-kins* are well described in the following quotation from Landa's "Relación de las cosas de Yucatan":

---

[3] Lopez de Cogolludo, 1867–1868, vol. 1, bk. 5, chaps. 16–19; Documentos para la historia de Yucatan, vol. 2, doc. 18; Scholes and Adams, 1938, vol. 1, introduction.

> They taught the sons of other priests, and the second sons of the chiefs, who were brought to them very young for this purpose, if they found them inclined toward this office.
>
> The sciences which they taught were the reckoning of the years, months, and days, the festivals and ceremonies, the administration of their sacraments, the omens of the days, their methods of divination and prophecies, events, remedies for sicknesses, antiquities, and the art of reading and writing by their letters and the characters wherewith they wrote, and by pictures that illustrated the writings.[5]

Landa also states:

> Some of the principal lords were learned in these sciences, from interest, and for the greater esteem they enjoyed thereby; yet they did not make use of them in public.[6]

The last phrase of the foregoing quotation would indicate a differentiation of sacerdotal and governmental functions, and the existence of a powerful priestly class tends to confirm this view. But during the investigation of idolatry made by Landa in 1562, testimony was received revealing the fact that caciques had actively participated in sacrificial rites. This may have been due, however, to the new conditions resulting from the Spanish conquest.

The clergy and the provincial authorities consciously labored to restrict the authority and undermine the influence of the caciques and *ah-kins*. One means to this end was the new system of education by which the sons of caciques, *principales*, and other native leaders were withdrawn from the care and teaching of their parents and taught Christian doctrine and ethics in mission schools. Provincial legislation was also formulated for the government of Indian affairs and the administration of the pueblos. The ordinances promulgated in 1552 or 1553 by Lic. Tomas López, judge of the Audiencia of Guatemala, during his *visita* of the province of Yucatan, covered every phase of local administration and included a number of sections designed to limit and control the authority of the caciques.[7] In one section López took note of the fact that the caciques were accustomed to hold secret meetings and to summon the common people to outlying districts in order to teach them the ancient rites,

---

[4] The term *principal* (chieftain) was generally used by the Spaniards to designate the Indian official who was in charge of a *barrio*, or subdivision, of an Indian town.

[5] Landa, 1937, p. 13.

[6] *Ibid.*

> . . . telling them that their former gods had given them messages concerning things that should be done, and making predictions of things that would happen if they were not done, and threatening the [common people] by other similar means in the name of their gods.

In order to combat this evil, it was decreed that henceforth

> . . . no Indian, man or woman, native of this said province, of any condition whatever, shall presume to make such summons and meetings, nor shall the said lords send *coyoles*[8] to any Indian, nor preach and teach, publicly or in secret, their rites and past idolatries or things pertaining to their gods, nor renew the memory of them, nor call meetings for the purpose of revolt and rebellion against the King in any way whatever.

The prohibitions and penalties imposed by the provincial authorities were supplemented by disciplinary measures imposed by the clergy in the missions. Once the Indians were converted, they were subject to the usual ecclesiastical pains and censures for heresy and other offenses against the Church and its teachings. In cases of major violations of the faith by the Indians, the clergy sometimes employed forms of corporal punishment.

In their efforts to combat the influence of the native leaders, the clergy and provincial authorities labored under a serious handicap, deriving from the system of local government introduced after the Conquest. Within the new system of government introduced by the Spaniards, the Indian pueblos continued to enjoy a considerable degree of local autonomy. For at least half a century after the Spanish occupation, the local officials of the pueblos were recruited largely from the former lords and chieftains, for the Spaniards, wherever possible, recognized the lordship and leadership of the former ruling families. For example, Nachi Cocom, *halach uinic*[9] of the former *cacicazgo*, or state, of Sotuta, named Juan Cocom after his baptism, became cacique of the pueblo of Sotuta. The head of the Xiu family, named Francisco de Montejo Xiu subsequent to his conversion, was made cacique of Mani, and several other members of the family were retained as governors of other pueblos in the Mani area. In the former *cacicazgo* of Hocaba-Homun, five brothers of the former ruling Iuit family became caciques of important pueblos. Examples could also be cited for other parts of the province. Moreover, it appears that in two cases, at least, heads of former reigning houses were permitted to retain a degree of authority and influence over larger areas roughly corresponding to the ancient states. This was true of Francisco de Montejo Xiu of Mani and Francisco Namo Iuit, cacique of Hocaba. Finally, there is evidence that certain *ah-kins* obtained local office as caciques or *principales*. In the course of time this situation was altered, and the proportion of former lords and members of old ruling families in the class of pueblo officials was sharply reduced, but in the sixteenth century, especially during the first fifty years after the Conquest, they exerted a powerful influence in pueblo affairs. Thus the system of local government tended to offset the various pressures and disciplinary measures designed to combat the loyalty of the Indians to their old religion.[10]

---

[7] Lopez de Cogolludo, 1867–1868, vol. 1, bk. 5, chaps. 16–19.

[8] *Coyol:* "prenda, señal que se da en los mercados," Perez, 1866–1877, p. 58. It is also defined as a sign given by the Indians when they wished to go to war. Scholes and Adams, 1938, vol. 1, doc. 10.

II

The contemporary documents contain brief references to cases of idolatry that were discovered during the decade from 1550 to 1560. Padre Lorenzo de Monteroso, who served as missionary in Sotuta for a few years, punished certain Indians as idolators. In 1558 Fray Diego de Landa, who then held the office of custodian (*custodio*) or prelate of the Franciscan Order, discovered "great evils and idolatries" during a visit to the Valladolid area. With the aid of the *alcalde mayor*[11] of the province, he called together the Indian lords of that region on the occasion of a religious festival and admonished them to repent and to abandon their idols. In 1559 a sacrifice was performed in Hocaba for which several Indians were punished, and in the following year Padre Rodrigo Muñoz, missionary in the Cochua province, punished certain Indians of the pueblo of Tepich because they possessed idols and had been guilty of idolatrous practices.[12]

---

[9] Head chief.

[10] Scholes and Adams, 1938, vol. 1, introduction.

[11] The *alcalde mayor* was the head of the Spanish provincial government.

In 1561 a newborn infant was apparently crucified in the pueblo of Hunacti. It was reported at the time that the child had been born dead "with the signs of crucifixion on the hands, feet, and side, and with the marks of thorns on the head." The body had been immediately buried, but by orders of Fray Pedro de Cuidad Rodrigo the corpse was disinterred and brought to Mani for examination. Because of rumors that "some evil thing" had been done, Fray Pedro publicly expressed the view that the marks on the body could have been "a natural thing," and a year later he reiterated this view and refused to admit that the child had been crucified. But in a memorial presented by Fray Diego de Landa in 1566 as part of his defense against charges resulting from the idolatry proceedings of 1562, to be described below, this incident is cited as a clear case of crucifixion and Fray Pedro's attitude is denounced as sheer stupidity.[13]

In the summer of 1562 Fray Diego de Landa, then provincial of the Franciscan Order, made a searching investigation of the practice of idolatry in the region which, in pre-Conquest times, had comprised the independent Maya states of Mani, Sotuta, and Hocaba-Homun. In the past our sources of information concerning this incident have consisted of a few contemporary letters and reports, the brief description given by Landa in his "Relación de las cosas de Yucatan," and the second-hand account in Fray Diego Lopez de Cogolludo's "Historia de Yucatan," written almost one hundred years after the investigation took place. Consequently the traditional version of the affair appearing in the histories of Yucatan has been incomplete and inaccurate. Recent research in the Archivo General de Indias in Seville has brought to light a mass of new or unused documents bearing on the case. These contain the official résumé of the proceedings sent to Spain by Bishop Toral in 1563, series of sworn declarations made by the Indians in 1562, hundreds of pages of testimony given by both Spaniards and Indians at a later date, and numerous miscellaneous letters, petitions, and reports.[14] On the basis of these papers and the materials already available it is now possible to reconstruct the story of the investigation and to form a more impartial judgment concerning its significance for the history of Yucatan and Spanish America.

---

[12] Scholes and Adams, 1938, vol. 1, introduction.

[13] *Ibid.*, vol. 1, doc. 20; vol. 2, doc. 82.

Toward the beginning of May 1562, two Indian boys of Mani discovered some idols and skeletons in a near-by cave. They immediately informed Fray Pedro de Cuidad Rodrigo, *guardián*[15] of the monastery of Mani, who gave orders to have the idols and skeletons brought to the patio of the monastery. Fray Pedro apparently transmitted news of the discovery to Fray Diego de Landa, who commissioned him to make an inquiry. At this time several other friars were residing in Mani, having been sent there to learn Maya, and with their aid Fray Pedro made a preliminary investigation. During the succeeding three or four weeks hundreds of Indians were arrested and questioned. Torture was used in order to obtain confessions. Many Indians confessed that they possessed idols and had performed various idolatrous rites, and as punishment for their offenses the friars had them whipped, imposed small fines, and forced them to take part in *autos de fe* that were held in the cemetery and patio of the monastery on three or four succeeding Sundays.

Convinced that the practice of idolatry was widespread, the friars sent letters to Merida urging Landa to come to Mani and take personal charge of the proceedings. Landa conferred with Dr. Diego Quijada, *alcalde mayor* of the province, who agreed to appoint an *alguacil*, or bailiff, with instructions to aid the provincial, arrest Indians accused of idolatry, and execute such sentences as he might impose. Soon after Landa arrived in Mani early in June, a definite plan of action was adopted for the continuation of the inquiry. It was decided that the provincial, together with Fray Pedro de Cuidad Rodrigo, Fray Miguel de la Puebla, and Fray Juan Pizarro, should remain in Mani to serve as judges of the Inquisition for the trial of the idolators, and that the other friars should go to the surrounding pueblos to inquire into the prevalence of idolatry, punish minor offenders, and send to Mani for formal trial and sentence those Indians who were guilty of serious offenses. In order to avoid delay and the duplication of legal costs, it was also agreed that full legal procedure should be omitted in the cases of minor offenders, especially the common people (*maceguales*), by combining them into one general process. The prosecution of major offenders, however, was to be in accordance with the customary forms of procedure.

---

[14] Extensive portions of these materials are being published in Scholes and Adams, 1938.

[15] The term *guardián* designated the head of a Franciscan monastery.

By what right did Landa assume authority to act as judge of Inquisition? Prior to 1571, when a tribunal of the Holy Office was established in Mexico City with jurisdiction over New Spain and Guatemala, the duty of trying cases of heresy and other offenses against the Church devolved on the bishops in their respective dioceses. But at this time there was no bishop in Yucatan, although Fray Francisco de Toral, a veteran of the Mexican missions, had already been appointed and consecrated bishop of the province and was then on his way from Spain.[16] By virtue of papal bulls,[17] however, the prelates of monastic orders in America in areas where there was no resident bishop had authority to exercise all episcopal functions except those that could be performed only by a consecrated bishop. Thus, from 1545 to 1562, the Franciscan prelates of Yucatan had possessed supreme ecclesiastical authority in Yucatan and had served as ecclesiastical judges ordinary in cases pertaining to the Church and the canon law. By a decree of February 28, 1558, the Audiencia of Guatemala had officially recognized and confirmed their authority.[18]

During the month of June and the first half of July the investigation moved forward rapidly. Hundreds of Indians were questioned. Many voluntarily confessed their guilt, others only after being subjected to torture. Landa and his associates soon became convinced that the caciques, *principales*, schoolmasters, and other prominent Indians were the chief offenders, and were responsible for the backsliding of the common people. Orders were given for the arrest of some forty Indians of the upper class, including ten governors and caciques of the Mani area, among whom was Francisco de Montejo Xiu.

The proceedings soon aroused increasing hostility on the part of the Indians, and Landa began to fear that they might resort to violence. He also desired the presence of the *alcalde mayor* at the time when he would pronounce sentences in all pending cases. Consequently, toward the end of June he sent word to Quijada asking him to come to Mani and to bring an escort of Spaniards, especially those who held *encomiendas*[19] in the Mani area. Quijada arrived early in the following month, and soon began actively to participate in the inquiry. With Landa's permission he questioned several Indians, even subjecting them to torture in order to force them to confess the full extent of their guilt.

---

[16] Toral was appointed bishop in 1560. After receiving notice of his nomination he went to Spain, where he was consecrated at Puerto Santa Maria on May 12, 1562. Documentos para la historia de Yucatan, vol. 2, p. iii.

[17] *Alias felicis* of Leo X, April 25, 1521; *Exponi nobis* of Adrian VI, May 10, 1522; and *Alias felicis* of Paul III, February 15, 1535. Hernaez, 1879, vol. 1, pp. 377–393.

[18] Scholes and Adams, 1938, vol. 1, doc. 9.

On July 11 the provincial pronounced sentence in most cases. The Indians guilty of minor offenses, especially the common people, were condemned to pay small fines, to participate in a general *auto de fe*, and, in certain cases, to be whipped. Certain major offenders, including several *principales* and schoolmasters, were condemned to receive one to two hundred lashes of the whip, to have their hair shorn (a grievous thing for an Indian), to pay heavier fines, to be deprived of local office, or to serve terms of forced labor for one to ten years. Inasmuch as the documents contain very little direct testimony given by the Indians at Mani, there is not much information available concerning the nature of the offenses for which these penalties were imposed. Most of the common people, however, were probably guilty of little more than mere possession of idols. Persons regarded as guilty of more serious offenses had apparently burned copal before the idols or had performed sacrificial rites, such as the offering of birds and deer or acts of self-mortification. Although we have very little evidence of human sacrifice in the Mani area, this is doubtless due to the marked lack of Indian testimony from Mani in the available documents. There is no reason to suppose that human sacrifices were performed less frequently in the Mani district than in other parts of the province.

On the following day, July 12, a public *auto de fe* was held in the pueblo of Mani. The sentences were read, and the *alcalde mayor*, as the representative of civil authority, agreed to their execution, stating that they were just and had been given with his approval and sanction. These formalities were proclaimed to the Indians by an interpreter, following which the sentences were executed. During the investigations, testimony had been received that several Indians who had died during the past year or two had also been guilty of idolatrous practices. By order of Landa the bones of these persons had been disinterred, and they were now burned as a part of the *auto de fe.* Thousands of idols that had been gathered up during the inquiry were also destroyed. The proceedings closed with a solemn mass of penance.

---

[19] The *encomienda* was a royal grant or favor made to Spanish conquerors as a reward for their services. At first it was essentially a grant of authority to exploit the labor of a stated number of Indians or those of a stated pueblo, but it later became merely a grant of tribute to be paid by the pueblo or pueblos mentioned in the *encomienda* grant.

The cases of twenty-five caciques, *principales*, and other major offenders were continued for further investigation. In view of their prominence as leaders of the native population, it was decided that they should be removed to Merida and held in confinement pending final action on the charges that had been filed against them. Included in this group were Francisco de Montejo Xiu, cacique of the pueblo of Mani, and seven other governors and caciques of the Mani area.

The friars now transferred their activities to the former *cacicazgos* of Sotuta and Hocaba-Homun. In these areas they followed the same procedure as at Mani, arresting large numbers of Indians, using torture in order to obtain confessions, and dispensing with full legal procedure except in the most important cases. Fray Diego de Landa directed the proceedings in Sotuta, whereas those in Hocaba-Homun were in the charge of Fray Andres de Bruselas, head of the monastery of Homun. Inasmuch as the hearings in Hocaba-Homun were oral, we have no record of the testimony given in that area at this time. At Sotuta, however, Landa received the sworn declarations of a large number of Indians from several pueblos in the Sotuta district, and these contain a mass of information concerning the performance of human sacrifices, as well as the offering of animals. As at Mani, the guilty Indians in these areas received various forms of punishment, such as fines, whipping, and terms of forced labor. And in each area the most prominent offenders, including numerous caciques and *principales*, were sent to Merida to be held in jail pending final sentence.

On August 14, 1562, Fray Francisco de Toral, first resident bishop of Yucatan, arrived in Merida. Landa's authority to serve as ecclesiastical judge came to an end and the bishop took charge.

This event marked a sharp change in the character of the inquiry, for the bishop immediately made it known that he disapproved the use of torture and severely censured the friars for the manner in which the investigation had been conducted.

The activities of Landa and his associates had aroused considerable unrest among the Indians. Indeed, according to Bishop Toral, the province "was on the point of being lost." The chief cause of unrest was the severe torture to which many of the natives had been subjected. The method most frequently employed was to suspend the Indians in mid-air by means of ropes tied to their wrists. If this proved ineffective, heavy stones were then tied to their feet. They were also whipped while in this position, and in many cases hot wax was dropped on their naked bodies. In certain cases the ancient form of torment by water was employed, as well as the *burro*, the latter being a wooden frame to which the Indians were tied and then subjected to other forms of punishment. So severe were these measures in some cases that they resulted in permanent injury or death. There is also evidence that a few Indians committed suicide—of whom the most prominent was Lorenzo Cocom, who had succeeded his brother Juan Cocom as cacique of Sotuta in 1561—in order to escape from the torment, or because of fear of the penalties that the inquisitors would impose as sentence for their idolatries.

The Indians also protested that much of the testimony concerning the practice of human sacrifice that had been given during the inquiry was false, that they had deliberately committed perjury in order to escape further torture. They protested that in certain instances they had agreed in advance to make false statements, or that when one of their number had given perjured testimony he told the others what he had said and the latter then repeated his story. Likewise, they asserted that many of the idols that they had handed over to the friars had been newly made for that purpose, that others were objects of worship long abandoned, and that in some cases they had gone to abandoned sites, in one case as far away as Coba, to gather up idols in order to meet the insistent demands of the friars.

It is clear that the situation was fraught with some danger. Prior to the bishop's arrival the *cabildo*[20] of Merida had sent two prominent citizens to Sotuta to beg Landa to use less drastic measures, but he refused their request. Faced by the intransigent attitude of the provincial, and realizing that protest to the *alcalde mayor* was futile, in view of the fact that he openly supported and approved the provincial's procedure, the representatives of the *cabildo* decided to make a secret appeal to the viceroy in Mexico City. The timely arrival of the bishop made this appeal unnecessary.[21]

---

[20] The local governing council.

Thus Bishop Toral faced a very serious problem. More than one hundred caciques and *principales* from Mani, Sotuta, and Hocaba-Homun were under arrest in Merida and it was rumored that Landa was of the opinion that many should be burned. Immediate action was necessary to quiet these fears and to ease the tension that was felt in all quarters.

After consultation with the *alcalde mayor*, the clergy, and a number of leading citizens, the bishop apparently decided to permit Landa and another Franciscan to continue the inquiry, temporarily at least, but on condition that the use of torture should be abandoned. Landa insisted, however, that the Indians would never confess the full extent of their guilt if not subjected to torment, and when the bishop refused to consent to such measures the Franciscans withdrew from further participation in the proceedings, except in the case of the reopened hearings in Hocaba-Homun to be noted below. Toral then assumed personal direction of the investigations. He decided (1) to question the caciques in person and to summon for reexamination many of the witnesses who had testified against them, and (2) to obtain written testimony in cases where it was not already available.

The caciques, when summoned before the bishop, protested that they were innocent of the charges of idolatry and sacrifice that had been filed against them, and several who had made confessions of guilt asserted that they had done so in order to escape torment. Likewise, many of the witnesses, on reexamination, repudiated their former declarations. Toral then sent a personal agent, Juan de Magaña, to the Sotuta area with instructions to check certain cases of sacrifice concerning which Landa had received written testimony prior to the bishop's arrival. After two weeks in the field, visiting several pueblos near Sotuta, Magaña reported (1) that the Indians had asserted that most of the persons alleged to have been sacrificed had never existed, and (2) that they had told him how many of the witnesses who had testified before Landa had made false statements in order to escape torture.[22]

---

[21] The above résumé of the idolatry inquiry as conducted by Fray Diego de Landa and his associates is based on Scholes and Adams, 1938, vol. 1, docs. 10, 12, 14, 15, 28, 31, 32.

Early in September Toral also commissioned Fray Andres de Bruselas and Fray Miguel de la Puebla, who had conducted oral hearings in the Hocaba-Homun region during the preceding July and August, to reopen the investigation in that area and obtain written testimony. He also instructed them not to employ torture under any circumstances. In accordance with these instructions they received sworn declarations from a number of Indians, all of which contained evidence of human and animal sacrifices that had been performed in the Hocaba-Homun area during the preceding ten years.[23]

Thus we face a very important question in dealing with the testimony concerning the practice of sacrificial rites. Was it true or false? As noted above, the Sotuta testimony was repudiated by the witnesses who gave it. On the other hand, we have the Hocaba-Homun evidence, presumably obtained without the use of torture, which confirmed the general character of the testimony from Sotuta. We cannot suppose that conditions in one area differed from those in the other.

Various explanations may be offered to resolve this problem. It is possible that in the Hocaba-Homun hearings conducted by Bruselas and his associate early in September, torture was actually used in violation of Toral's orders. The documents contain no hint that this was true, but inasmuch as these proceedings constitute the final share of the Franciscans in the general idolatry investigation, it may be possible that Bruselas employed torture and that this phase of the inquiry marks the point where the final break came between the bishop and the friars. If this is true, then the Hocaba-Homun evidence has no greater validity than that obtained by Landa at Sotuta.

The crucial point is how to regard the Sotuta testimony. In the first place, it should be noted that when the bishop questioned the caciques and *principales* in Merida, he assured them that he would give them a fair hearing and do justice in the cases pending against them. It would not be surprising, therefore, if the Indians took advantage of his attitude to minimize their guilt. Second, there is evidence that news of the bishop's openly expressed censure of Landa's conduct, especially in the matter of torture, soon spread far and wide among the Indians, and they may have interpreted this as condoning, if not actually approving, their idolatrous practices. The friars asserted that this was true and they even submitted some documentary proof for their point of view.[24] It may be argued, however, that these considerations should have prompted the Indians of Hocaba-Homun, assuming that they were not put to torture, to deny the existence of idolatry when Fray Andres de Bruselas and Fray Miguel de la Puebla received written testimony from them early in September. But this argument may be offset by taking into account the time element. Toral's questioning of the caciques and *principales* coincided in time with the hearings in Hocaba-Homun, and the Indians of the latter area were probably unaware of the bishop's attitude. But later in the month, when the bishop sent Magaña to Sotuta, his views were well known.

---

[22] Scholes and Adams, 1938, vol. 1, doc. 28.

[23] *Ibid.*, vol. 1, doc. 14.

There is also reason to believe that Magaña's inquiry in the Sotuta area was not very searching. Landa later characterized it as "a thing to mock and sneer at, because there is not an Indian, man or woman, who would dare to accuse a lord or chieftain, even though they were burned alive." A careful reading of Magaña's report also leaves the impression that the Indians were far more anxious to tell how witnesses who had appeared before Landa had given false testimony than to answer questions about alleged cases of sacrifice. Moreover, Magaña wrote a letter to a friend in Mexico City in which he expressed the view that the Indians would naturally repudiate their former testimony when summoned before a new judge, referring, of course, to Bishop Toral.[25]

Analysis of the testimony from both Sotuta and Hocaba-Homun provides the best means of resolving the doubt concerning its validity. Such an analysis reveals that witnesses who described a given case of sacrifice were in agreement on many of the details, such as time, place, and participants. It must be admitted that there

[24] *Ibid.*, vol. 1, doc. 18.

[25] *Ibid.*, vol. 1, docs. 22, 28; vol. 2, doc. 81.

were certain discrepancies in the testimony concerning the same case, but this may have been due to faulty memory, inasmuch as many of the sacrifices had been performed one, two, and, in some instances, ten years prior to the time the testimony was given. It can be argued that agreement concerning details may have been due to consciously planned perjury on the part of the Indians, but if this is true then we must find some other explanation for the discrepancies.

Another powerful argument in favor of the general validity of the testimony is the fact that it confirms, supplements, and clarifies our knowledge of Maya religion derived from other sources. Moreover, the descriptions of the sacrifices contain elements that must surely have been due to Christian influence and new conditions resulting from the Conquest. Indeed, we find so many references to sacrifices performed inside the churches, or to the mixture of Christian and pagan customs, that the descriptions must refer to post-Conquest cases.

We are convinced, therefore, that a large part of the evidence must be accepted as true. It is possible that it contains exaggerations, even certain falsehoods, and that the Indians may have called upon memory of pre-Conquest practices as regards certain details. But these considerations would hardly seem to warrant the assumption that the testimony does not provide a generally valid description of actual cases of idolatry and sacrifice that had occurred subsequent to the Spanish occupation and the beginning of the missionary program.

In his public and official statements Bishop Toral adopted the point of view that the testimony, at least the Sotuta series, contained falsehoods and exaggerations. There can be no doubt, however, that he recognized the existence of idolatrous practices, although minimizing their importance and significance; but he believed that the Indians should not have been subjected to torture and rigorous punishment. On several occasions he roundly denounced Landa and the friars for having employed such stern measures against Indians "so new in the faith," and expressed indignation that they should have sustained the view "that the law of God could not be preached without torments." Moreover, he was undoubtedly impressed by the fact that the situation when he arrived was extremely delicate and that expediency demanded a drastic change in policy.

Finally, it must also be recognized that personal animosity and provincial rivalries soon became factors in the situation. The *alcalde mayor's* administration of provincial affairs had already aroused great discontent. Citizens who had reason to complain of his conduct now denounced his support of Fray Diego de Landa, especially the manner in which he had given the friars the aid of the secular arm during the Mani inquiry, as both unjustifiable and illegal. Responsibility for the torture of the Indians and for the deaths that had resulted therefrom was now laid at his door, and when his term of office came to an end in 1565, charges of this kind were filed against him during the *residencia* conducted by his successor. It was undoubtedly true, as the Franciscans asserted, that Bishop Toral leaned heavily on the discontented factions for advice and counsel, and that he was unduly influenced by their views. The friars also denounced the bishop's attitude as constituting disloyalty and treason to the Franciscan Order, the bishop being a Franciscan. Such criticism probably cut Toral to the quick, and in the course of time personal feeling may have influenced his decisions.

The bishop's refusal to sanction the continued use of torture, his general attitude of moderation, even partiality, in dealing with the cases of the caciques and chieftains imprisoned in Merida, and his decision not to extend the judicial hearings to other parts of the province, as the Franciscans proposed, decreased the tension and unrest among the native population. He found it possible, therefore, to proceed slowly with the cases of the caciques and *principales*, in accordance with customary legal formulæ, and it was not until January 2, 1563, that he announced the final sentences. Unfortunately, a copy of the sentences has not been preserved in the documentary materials relating to the inquiry, but from other sources we gather that mild forms of physical punishment were imposed, as well as certain ecclesiastical censures. In reports to the king, the bishop stated that he believed the sentences to be just, "taking into account the long imprisonment [the Indians] had endured, as well as the banishment from their pueblos and provinces [and] loss of property, and that more severe punishment therefore was not advisable."[26]

The failure to acquit the Indians of the charges of idolatry and the justification cited for the light penalties that were imposed sustain the view expressed above, that Bishop Toral was aware of the existence of idolatry. His decisions in these cases were based on expediency and a genuine desire to see justice done. The imposition of rigorous punishment on one hundred and twenty-three caciques, *principales*, and other native leaders, several of whom were former rulers of some prominence, would have caused tremendous unrest, if not violence, on the part of the Maya. But it was also Toral's sincere view that it was contrary to the spirit of Christian charity and justice to subject the Indians, so recently converted, to penalties of a drastic kind. The same considerations prompted his decision, made a month later, to revoke the sentences of all the Indians whom Landa had condemned to terms of forced labor.[27]

---

[26] Scholes and Adams, 1938, vol. 1, doc. 28.

By this time Landa and Toral had become hopelessly alienated, and each now turned his attention to the preparation of reports to be sent to Spain. At the bishop's request witnesses were summoned to testify concerning certain alleged usurpations of authority by Landa and other Franciscan prelates in earlier years, the careless, if not improper, manner in which the Franciscans, in Toral's opinion, had conducted the missions, and the proceedings of the provincial and his associates during the idolatry inquiry. This report and an incomplete and somewhat circumstantial résumé of the whole investigation were sent to Spain in March 1563. They were accompanied by two covering letters in which the bishop denounced Landa and the friars in scathing terms.[28] Landa likewise prepared reports of a justificatory character, for the purpose of proving that the friars had performed great services in Yucatan, that his conduct of the idolatry proceedings had not been characterized by the severity and lack of justice that the bishop alleged, and that the prevalence of idolatry and human sacrifice thoroughly warranted the action that he had taken.[29] In order to present his case before the king and Council of the Indies in person, he resigned as provincial of the Franciscan Order, and sometime during the spring of 1563 he departed for Spain. The reports prepared by Landa and the bishop were also supplemented by letters despatched by Dr. Diego Quijada, the *alcalde mayor*, the

[27] Scholes and Adams, 1938, vol. 1, doc. 30.

[28] *Ibid.*, vol. 1, doc. 31; vol. 2, docs. 41, 43.

[29] *Ibid.*, vol. 1, doc. 32.

cabildo of Merida, the officials of the royal treasury, and other persons.[30]

The bishop's reports and letters were received in Spain early in 1564 and they made a profound impression. Consequently, when Landa arrived later in the year, having been delayed by illness and other circumstances, he found the members of the Council of the Indies in a hostile frame of mind. But with the aid of friends, he was permitted to present his version of the case and to file with the Council numerous papers, including copies of part of the Indian testimony concerning human sacrifices, in justification of his actions. One of the major questions involved was whether Landa had legal authority in 1562 to serve as inquisitor and ecclesiastical judge. Because of this aspect of the case and the wider implications that would follow if the Council imposed discipline on a prominent ecclesiastic, as well as the fact that many of the accusations that Toral had made related to the internal affairs of the Franciscan Order, it was decided to remit the entire matter to the provincial of the Franciscan province of Castile for review and decision, with authority to impose upon Landa whatever punishment he might deem necessary.[31]

During the spring of 1565 Landa presented an elaborate defense before the person appointed by the provincial of Castile to review the case. Inasmuch as the crucial question was whether Landa, as provincial of the Franciscan Order in 1562 prior to the arrival of Bishop Toral, had the right to serve as ecclesiastical judge, legal opinions on this point were sought from a number of civil and canon law jurists, of whom one was Lic. Tomas López, who had governed Yucatan for a year (1552–1553) by virtue of a commission from the Audiencia of Guatemala. These legal specialists all agreed that various papal bulls and colonial administrative decrees warranted Landa's exercise of jurisdiction and that the penalties he had imposed had not been excessive.[32]

The filing of these opinions assured Landa's acquittal, although owing to the death of the provincial of Castile, a decision was not rendered until four years later. On January 29, 1569, the provincial's successor made a formal pronouncement to the effect (1)

[30] Cartas de Indias, pp. 380–396; Scholes and Adams, 1938, vol. 2, docs. 34–40, 42, 44–45.

[31] Scholes and Adams, 1938, vol. 2, doc. 78.

[32] *Ibid.*, vol. 2, doc. 83.

that Landa had not exceeded his authority, and (2) that it had been his duty to act as inquisitor and punish the Indians guilty of idolatry, invoking the aid of the secular arm. It was his decision, therefore, that Landa was without guilt and was free of all the charges on which he had been accused.[33] Landa's vindication was complete when, three years later, Philip II nominated him as bishop of Yucatan to succeed Fray Francisco de Toral.

In so far as the decision of the provincial of Castile was based on the legal questions involved, it was entirely just. We believe, however, that Landa deserved severe condemnation for the manner in which he employed torture during the idolatry inquiry, and it may be doubted whether he escaped private censure by his superiors on this point. Landa was convinced, and justly so, that the province of Yucatan was rotten with idolatry, and it was clearly his duty, as the governing prelate, to find a remedy. Human sacrifice was a horrible thing that had to be stamped out if Christian standards of civilized life were to be respected, and Landa's activities undoubtedly contributed to that end. Moreover, we must not forget that the use of torture was a recognized legal instrument in the sixteenth century and that it was occasionally employed in idolatry proceedings in other parts of America during this period. Nevertheless, we have more sympathy for Bishop Toral's view that from the standpoint of the missionary program the extreme measures employed by Landa were neither humane nor expedient.

III

The testimony concerning idolatry and human sacrifice that was received during the proceedings of 1562 deserves detailed discussion and analysis. No other set of documents at present available provides such significant information concerning the practice of Maya religion subsequent to the Spanish occupation of Yucatan and the beginnings of that process of fusing the Maya and Christian religions which has continued down to the present day. The evidence also contains invaluable data which clarify and enlarge our knowledge of Maya religious customs prior to the Conquest.[34]

---

[33] Scholes and Adams, 1938, vol. 2, doc. 85.

[34] The summary of the idolatry testimony, as presented in section III, is based on the materials in Scholes and Adams, 1938. Citations to these materials will not be made in this section of the paper except in the case of direct quotations.

It has been so often stated that Maya science was lost largely because of the extermination or driving out of the native priests after the Spanish conquest that it is of special interest to note in the documents of the 1562 proceedings that a large number of these priests were still living and secretly practicing their profession. Only once local priests are named as taking part in human sacrifices at the pueblo of Sotuta, but we find three assisting priests from Yaxcaba and Tixcacaltuyu also taking part. At Kanchunup six priests are named, at Mopila five, Yaxcaba four, Tibolon four, Usil two, and at Sahcaba only one. All but two at Mopila had Christian names and had evidently been baptized; and one of the Usil priests, Juan Pech, had learned his profession while acting as schoolmaster, an office which he still held. Pressure by the missionaries no doubt diminished the number of neophytes as time went on; but it now seems probable that the disappearance of Maya religious and scientific lore was very gradual and Maya astronomy in course of time gave way to European astrology, of which we find much in the later Books of Chilam Balam. From the 1562 inquiry we learn that the local schoolmasters were frequently present at human sacrifices; and reports made by the *encomenderos* to the king in 1579 still accuse these men of being great idolators and even keeping idols in the schoolhouses.[35] There can be little doubt that this class continued to carry on the old traditions for a long time and were the men who compiled the so-called Books of Chilam Balam and kept them in circulation.[36]

Conversion to Christianity, of a sort, appears to have been quite sincere among the natives generally, and at first even among some of the priesthood,[37] who saw in the coming of the Spaniards the prophesied return of Kukulcan with his white-robed priests; but it was not the exclusive faith taught by the missionaries. The Maya very soon attempted to adapt the new teaching to their own religious pattern and superimpose it upon their old beliefs. While most of the native sermons reported in the present inquiry utterly condemned the friars and their doctrine, at Tibolon the officiating *ah-kins* told their hearers

---

[35] Relaciones de Yucatan, vol. 2, pp. 190, 213; Roys, 1933, pp. 4–5.

[36] Sanchez de Aguilar, 1892, p. 115.

[37] Ah Kin Euan of Caucel, afterward cacique of that town, was always a strong supporter of the missionaries. Lopez de Cogolludo, 1867–1868, vol. 1, bk. 5. chap. 6; Roys, 1933, p. 146.

> . . . that they had made sacrifice to their gods in order that they should give them food, provide them with what was necessary, and grant them rain for their milpas;

and they also said,

> . . . we should receive what the padres say to us, but we ought not on that account cease to perform our idolatries.[38]

This religious tendency has long been noted,[39] and we find striking additional evidence in the practice, as reported in these documents, of crucifying sacrificial victims, who were identified with the Christian Deity.

To the Maya, this superimposition of a new faith upon the old was no new thing. Already, centuries before, the Quetzalcoatl cult had been brought into Yucatan, notably at Chichen Itza, by Mexican invaders, who were also said to have introduced idolatry.[40] This movement, with its abolition of the old stela cult and the other religious changes it brought, could hardly have been to the advantage of the older Maya priesthood; but the new religion was reconciled with the old, and in course of time the descendants of the foreign conquerors were either driven out or merged with the native population,[41] although the ruling class continued to allege its Mexican descent.[42]

These things, it would seem, influenced the attitude of the natives toward the introduction of Christianity, and the Spaniards were compared with the Itza. In a prophecy for Katun 2 Ahau we read:

> Then shall be known the time of the white men, the bearded ones, as the priest Chilam declared, at the time of the arrival of their elder brothers, just as the passing Itza arrived among them. You shall intermarry with them, you shall dress in their garments, you shall put on their hats, you shall speak their language.[43]

But the Itza finally left the country. This may have been in the mind of the Yaxcaba priest, Gaspar Chim, who told the people, after he had sacrificed a dog in the church,

---

[38] Scholes and Adams, 1938, vol. 1, doc. 12.

[39] Thompson, 1930, p. 56; Roys, 1933, appendix G.

[40] Relaciones de Yucatan, vol. 1, pp. 121, 215, 226, 271.

[41] *Ibid.*, vol. 1, pp. 197, 204–205.

[42] Roys, 1933, appendix E.

[43] Book of Chilam Balam of Tizimin, p. 36; Perez Codex, p. 155.

> . . . that what the padres told them and preached to them was not the truth and they should not receive it; that it was a thing to jest at, and he whom they preached was not God. Nor would they continue in the land, and they must very soon go away; and they themselves and the other natives would return to their old things and would perform their ceremonies.[44]

Landa's drastic measures probably had a part in causing this intransigent attitude to disappear. Later Maya literature indicates a genuine feeling that while the old things should not be forgotten, Christianity had become a permanent institution.[45] Although Landa's efforts did not put an end to compromises between Christianity and paganism, there can be little doubt that human sacrifice was abandoned, at least among the people under Spanish domination. Idolatry died out more slowly; it is difficult to determine when it ceased. Seler tells us of the survival of idolatry in various parts of Mexico,[46] but in modern Yucatan we find no mention of idols in connection with such pagan deities as are still worshipped.

Landa tells us that the priesthood took the initiative in ordering sacrifices.[47] The 1562 inquiry indicates, however, that the caciques or *halach uinic* were usually the persons who commanded the more important sacrifices to be made. Moreover, in certain cases they actually assisted in the act of excising the hearts of the victims. It is possible, of course, that there was a change due to the Conquest.

Now that the old temples had been destroyed or abandoned, in the inhabited towns at least, it is natural to find sacrifices often taking place in the village church. Others were made in the churchyard, corresponding to those formerly solemnized in the temple patio.[48] At other times sacrifices were performed beside one of the crosses or heaps of stones at the entrance to the town, like the New Year ceremonies described by Landa.[48a] In 1559 offerings of food and drink were made in the church and at the crosses outside in order to bring rains. The church was adorned with leaves of *copo* (*Ficus cotinifolia* H. B. K.) and the entire place was illuminated with candles and other lights. These ceremonies continued for seven days. As this sacrifice was said to have been made "in the ancient manner," we evidently have here an old temple festival adapted to the new condition.

---

[44] Scholes and Adams, 1938, vol. 1, doc. 12.

[45] Roys, 1933, p. 98.

[46] Seler, 1902–1923, vol. 2, pp. 87–93; vol. 3, pp. 363–364.

[47] Landa, 1928–1929, vol. 1, p. 198.

[48] *Ibid.*, vol. 1, p. 200.

[48a] *Ibid.* vol. 2, p. 18.

The house of the cacique was also the scene of these rites in one or two cases. In the Hocaba-Homun area practically all the sacrifices took place in isolated *milpas*, but this may have been due to the fact that there had been prosecutions in the past for sacrificing in the village churches.

A number of children were taken to Chichen Itza and thrown into the Sacred Cenote. A few were cast into the water alive to consult the "oracle," but most of them were first sacrificed to local idols and their bodies cast into the *cenote*. One case is described with certain interesting details. The Iuit caciques of Hocaba-Homun decided to send a boy of ten years to Chichen Itza to be thrown into the *cenote* "in order that he might speak to their gods." The messengers who executed this commission remained for a day after the boy had been cast into the water, waiting to see if he would reappear with the reply of the gods, "whether they were going to have some pestilence, good harvest, or drought."

Landa intimates that human sacrifice was resorted to only on special occasions, "por alguna tribulación o necessidad."[49] Such motives, according to the 1562 inquiry, seem to have been anxiety for rain and favorable weather for the crops, a great hurricane, and the failing health of the *halach uinic* of Sotuta. On one occasion at Sahcaba four idols, probably those of the rain gods, were brought to the church, and after an offering of Spanish wine mixed with honey and water, the wells were covered, in imitation of a fast, in order that the gods should provide rain for the *milpas*. At the time of a great hurricane, the blood, body, and heart of a human victim were thrown into a *cenote*. Here four idols, possibly wind gods, were worshiped, which suggests that the wind gods as well as the rain gods were associated with the *cenote* cult. At Sotuta on the occasion of the last illness of the *halach uinic* Juan Cocom (Nachi Cocom), he himself attended a human sacrifice, and later, when he was at the point of death, two more boys were sacrificed in the *cenote*.

The documents contain the names of several deities. Itzamna, the principal deity and perhaps the sky god, naturally appears. We also encounter a god of the hunt called "Aquehe," which can be safely reconstructed as Ah Ceh, although we know no deity of that name. Defined in the Motul dictionary as "hunter" or "archer," it could also mean "the principal male deer." As such it might well be identified with the Zip, cited by Landa as a deity of the chase, and still worshiped by the modern Indians. It takes the form of a small deer and is the guardian of these animals. *Zip* means "to miss one's shot," and unless Zip is suitably propitiated it will cause the hunter to miss his prey.[50] Another god, the "diablo Tabay," is not characterized in the idolatry documents. This was one of the five principal gods of the Acalans and was probably the patron of one of the four quarters of the city of Itzamkanac.[51] Landa identifies it as a hunting god;[52] and the name, which means "deceiver," could well be associated with that of the Zip. The documents also mention a deity called Kakalku ("fire," or "fiery god"). We know of no such god, but the name reminds us of Kak-u-pacat (*visto de fuego*), Al-chuy-kak (*el dios de sus batallas*), and Kinich Kakmo (*sol en rostro*).[53]

---

[49] Landa, 1928–1929, vol. 1, p. 198.

Landa tells us of the importance of genealogy among the Maya and the pride they took in their descent.[54] We now learn that this was a religious as well as a temporal matter, and that this lineage cult was one of the subjects commemorated in their hieroglyphic books. "Hunixquinchac" (Hun-ix-kin-chac?) was said to be the god of the people named Puc, and "Chocunquinchac" was the deity of the Kumun lineage. "The greatest of these gods" was Zacalpuc, with whom we are more familiar; but unfortunately we are not told with which lineage he was associated. We know him as one of the early Mexican conquerors of Yucatan and as head of one of the four lineages which came from heaven.[55] Indeed, he still figures in the Maya prayers of modern *yerbateros* as the first man to offer *posole* to the Chacs.[55a] We cannot but suspect that the ritual of the *holpops* in the first chapter of the Chumayel manuscript is dedicated to these family or lineage deities.

---

[50] *Ibid.*, vol. 2, p. 54; Redfield and Villa, 1934, pp. 117–118.

[51] Paxbolon-Maldonado papers. Archivo General de Indias, Mexico 138.

[52] Landa, 1928–1929, vol. 2, p. 54. We have read this passage as "Zip i Tabai" instead of the usual reading "Zipitabai."

[53] Lopez de Cogolludo, 1867–1868, vol. 1, bk. 4, chap. 2.

[54] Landa, 1928–1929, vol. 1, p. 168.

[55] Valladolid lawsuit of 1618, *in* Brinton, 1882, p. 115; Roys, 1933, pp. 53, 147.

[55a] Maya prayer dictated to R. L. Roys by Benito Cauich of Piste.

It is interesting to note that there is little mention of stone idols in the documents of the 1562 proceedings. Most of them were of pottery and a lesser number of wood. At one human sacrifice a large number of idols were brought, which belonged to the cacique of Huhi. Forty of them were of clay and two of wood. The last two were set upon small stools and covered or draped with foliage. The idols were both large and small, and apparently represented animal and human figures.[56] Their mouths, which were frequently smeared with blood during the sacrificial rites, are referred to as snouts (*hocicos*). The last feature corresponds to the representation of a number of the deities in the codices, including what are believed to be the rain gods.

Participants in the sacrifices apparently brought their own idols. The total number that was used in these ceremonies ranged from two to as many as sixty, the number frequently specified being four. The latter suggests strongly the gods of the four world quarters, such as the Chacs, the Bacabs, and the Pauahtuns. Of these the Chacs, or rain gods, probably received the most attention. According to Landa, the four Bacabs held up the sky,[57] but both they and the Pauahtuns have been reported as wind gods. We may well conclude that in times when the worship of the old gods was attended with so much danger, only those who were most important to the welfare of the people would receive sacrifices. These may have been the rain and wind gods and the tutelary divinities of the lineages.

In certain cases the offerings consisted solely of food and drink. The food was probably the native bread (*uah*) prepared in various ways and resembling the ceremonial breads still used in the *milpa* ceremonials. The modern *tuti-uah* is a maize bread mixed with ground squash seeds.[58] The drink was usually *posole*, which was sometimes mixed with chocolate. *Balche* was also sometimes drunk during the banquets which followed the various sacrifices and we may infer that it was part of the offering.

Another type of sacrifice was that of animals. The dog, deer, peccary, and turtle are specially mentioned among others. The documents specifically state that the hearts of the dogs and the deer were excised and burned as offerings before the idols, and it seems probable that the same was true of the peccary. Frequently a ceremonial fire was kindled from slender rods or withes and later extinguished with the blood of the offering. In the case of the peccary, the fire was put into the open carcass of the animal and then extinguished with blood. Crosses, sometimes taken from the church altar, were burned in these fires, but this was probably due to Christian influence. We doubt that it was, as alleged, done in disrespect to the cross. This ceremony, of course, was a variation of the fire-extinguishing ritual (*tupp-kak*) described by Landa,[59] although in the latter the hearts of animals were burned in the fire, which was then extinguished with water.

---

[56] One witness of the 1562 inquiry testified that the idols had "faces of women and demons and deer and other figures." Scholes and Adams, 1938, vol. 1, doc. 32.

[57] Landa, 1928–1929, vol. 2, p. 14; Lopez de Cogolludo, 1867–1868, vol. 1, bk. 4, chap. 8; Mendez *in* Saville, 1921, p. 176; Roys, 1933, appendix A.

[58] Redfield and Villa, 1934, p. 70.

The carcass of a sacrificed dog was afterward roasted and eaten, and the same was probably done with other animals.

The most important of the sacificial rites was human sacrifice. Captives of war were no longer available, and the victims were now all children. These varied in age from two years to ten or twelve. Two of the victims were mere babies who had just started to walk. Both boys and girls were immolated. No adult victims are mentioned in the testimony, though we know that many adult slaves, as well as war prisoners, were sacrificed in pagan times. It is very possible that even in the pre-Spanish period the majority of the victims were children.[60]

These children were secured for sacrifice in various ways. Many were kidnaped, some from neighboring towns, but more from outside the *cacicazgo* where the sacrifice occurred. A few had been stolen as far away as Chemax to the east, Sucopo and Loche beyond Tizimin, and Kinlacam in the region of Calkini to the extreme west. Others had been taken from the neighboring *cacicazgos* of Ceh Pech, Ah Kin Chel, and Mani, but the greater number probably came from Cochuah and Cupul. Although the Spanish chroniclers tell us nothing of this practice of kidnaping children, the fact that the documents state that it was the office (*oficio*) of certain men to obtain them in this manner suggests that it was of long standing.

---

[59] Landa, 1928–1929, vol. 2, p. 78.

[60] Thompson reports a high proportion of infant burials under temple floors at San José. He is convinced that at least some were sacrificed. The discovery of a fine jade amulet (jade is rare at San José) with one skeleton serves as additional evidence. In Mexico also many children were sacrificed to the rain gods, or Tlalocs. Sahagun, 1938, vol. 1, p. 84.

Other victims were purchased, both within and without the *cacicazgos.* The prices paid ranged from five to ten beads per head, probably red beads called *kan,* and not the more costly green beads called *tun.* In two cases, at Pustunich and the unidentified town of Quinçucche (Kinsucche?), a fathom of thick beads was paid for each of two small boys who were brought to Sahcaba near Sotuta to be sacrificed. Such children were usually the offspring of slave women. Landa tells us that in former times the entire town contributed to raise funds for such purchases.[61]

A witness from Mopila testified that one of the two men who had the office of kidnaping children for sacrifice in that area was an *ah-kin.* Certain boys sacrificed in Kanchunup had been purchased in the Cupul area by two *ah-kins.*

We are also told by Landa that pious persons offered their own children for immolation,[62] but according to the evidence presented in the 1562 inquiry such benefactors were men of means who contributed orphans living in their homes. The latter often bore the same patronymic as the donor, so they were evidently relatives. In one case the victim was a youth from Tekax, who had come to visit relatives at Yaxcaba. He was the only one from whom a protest is recorded. The priest attempted to reassure him, telling him he would go straight to heaven,[63] but the victim, who was a Christian, remained unconvinced.

One case is of especial interest. Francisco Chel, cacique of Cansahacab in the province of Ah Kin Chel, sent six little girls to the *halach uinic* of Hocaba to be sacrificed. The witness stated that a gift of this sort from one cacique to another was an ancient custom and called "*quimchich.*"[64]

Many names of victims are given. Some were Christians, but the greater number bore pagan names, indicating the children of parents who were as yet unconverted. Such cases would be less likely to come to the notice of the missionaries; but evidently a considerable number of the *maceguales* and slaves still remained unconverted at this time.

One human sacrifice at Yaxcaba is described in detail. On this occasion ten idols and other ceremonial paraphernalia were brought to the church from the *milpa* of the cacique. The idols were set in a row upon a bed of leaves of the *higuera de infierno*.[65] Before them a long mat was spread, and on it was laid a large flint knife, its *rabo* (handle?) wrapped in a white cloth. The victim, the youthful visitor from Tekax already mentioned, was stripped to the waist, blindfolded with a white cloth, and seated between two priests who sat on small stools. Later he was stretched on his back on the mat and his heart excised by a man whose office it had been to sacrifice people in former times. This was evidently the official *nacom*. Landa states that this office was "poco onrosa,"[66] but we doubt this, since it is shown in the present inquiry that *principales* and even caciques performed this service. The executioner raised the heart aloft, made two cross-like cuts in the point, and placed it in the mouth of the largest idol, called Itzamna.

---

[61] Landa, 1928–1929, vol. 1, p. 198.

[62] *Ibid.*

[63] It is interesting to find this confirmation of Thompson's conjecture "based on the known beliefs of the Aztecs and other Mexican peoples." Gann and Thompson, 1931, p. 130.

[64] This might be reconstructed as *cim ch'ich'* and tentatively translated as "kill the little ones," or "kill the birds."

There is little, if any, mention of a sacrificial stone in the testimony from the Sotuta area; but the more usual procedure was probably that described in the Hocaba inquiry. The victim was placed over a convex stone so as to raise the breast, where he was held by several persons, evidently the same as the *chacs* described by Landa.[67] The executioner excised the heart with a stone knife called *u kab ku* (the arm or hand of god), and gave it to the priests. One of the latter collected the blood and anointed the faces of the idols. Another placed the heart in what was evidently a new plate or flat bowl, called *zuhuy lac* (virgin plate) in Maya. It was covered by another similar plate[68] and offered by the priest to each idol in turn with a speech of presentation or invocation. After that the heart was frequently burned, it is said, in the same receptacle (*cajete*).

Copal was burned in a brazier (*cajete*) at practically all the ceremonies. At Hocaba we find an interesting detail, where the cacique, Francisco Iuit, fanned the burning incense with his mantle. It would seem that this was a ceremonial and not a casual action, for we read in the Interrogation of the Chiefs:

> [My] son, go and get me your loin-cloth that I may perceive its odor here with the wide spread odor of my loin-cloth, the odor of my mantle, the odor of my

---

[65] This may be the *kochle* or *Cecropia obtusa* Trec. Cf. Roys, 1931, p. 256.

[66] Landa, 1928–1929, vol. 1, p 194. His office was for life. There was another kind of *nacom*, the war-chief, who was chosen for three years.

[67] *Ibid.*, vol. 1, p. 200.

[68] Offerings between two flat bowls have been found in the ruins at Uaxactun.

censer. . . . This is the odor of the loin-cloth which he asks for, . . . it is copal gum set on fire [so that] it burns.[69]

On one occasion the gum of a rubber tree was burned with copal.

It is possible certain details of some sacrifices were supposed to be secret. Once in the Hocaba area the spectators of the actual sacrifice, ten of them *principales*, were dismissed before disposal of the heart was made; and only the caciques remained present. An unusual feature of this ceremony was the presence of more than twenty wooden masks adorned with branches, in addition to thirty or forty small clay idols.

This is of especial interest, because the Spanish reports tell us little, if anything, of the use of masks. They are not unfamiliar objects in the frescos and bas-reliefs; and the word for mask, *koh*, is defined in the Maya dictionaries, where it means both "mask" and "representative." Even in the San Francisco Dictionary, which dates from the latter half of the seventeenth century, we find the phrase "*kohin a koh . . . pon tu máscara*." Here, however, the lexicographer may have some European pageant in mind. Diego Chuc of Hocaba, formerly of Ixil, confessed to the ownership of idols, books, and masks and displayed four masks and some books at the inquiry.

Three unusual methods of sacrifice are related in the Hocaba inquiry. One is decapitation of the victim, before the heart was excised, offered to four idols, and burned. The body was afterward wrapped in a mantle, weighted with stones, and cast into the *cenote* at Bolonnicxan. Here we are reminded of the picture on page 34 of the Dresden Codex, where we see the severed head of the maize god over a glyph on an altar. Around the altar are several male figures playing musical instruments, and some objects, among which is a ladder.

We are also reminded of the same picture by a sacrifice at Bolonnicxan, in which two boys were tied to a ladder with arms outstretched, and then their hearts were excised in the usual manner.

A third and hitherto unheard-of form of sacrifice was almost unspeakably brutal. At the unidentified site of Yaxleuea two little girls and a boy were sacrificed to ten clay idols. The boy and one girl were immolated on the sacrificial stone in the usual manner.

---

[69] Roys, 1933, p. 94.

The other little girl was attached to a stake with her hands tied behind her. Then Juan Iuit, cacique of Hoctun and brother of the *halach uinic*, beat her over the breast with a thorny club of *pochote* wood (*Ceiba æsculifolia* Britt. & Baler) until she died. The bodies of all three were thrown into a deep dry cave, the mouth of which was covered with a large stone. We do not believe that this was merely to conceal the bodies, as in some cases they were simply left in the thick forest. It would seem rather that we have here the dramatization of some mythological legend, which has not come down to us.

In a sacrifice performed at Oxkutzcab a boy was tied to a stake, and while in that position his heart was excised.

The various manners in which the bodies were disposed of suggests that this depended upon the character of the sacrifice or the gods to whom it was made. In one case two orphan victims were buried in the church at Sahcaba near Sotuta. Whether the ancients made burials of victims beneath the temple, the writers do not know; certainly human remains have been uncovered in the excavations of platforms. More frequently they were buried in the cemetery by the church, which would correspond to either Christian or pagan custom; for Landa tells us victims were frequently buried in the temple court.[70]

A frequent method of disposal of the corpse of the sacrificed victim, which is not mentioned by the Spanish narrators, was to cast it into one of the local *cenotes*. The custom of throwing live victims into the *cenote* at Chichen Itza is, of course, well known;[71] but it now appears that the *cenote* cult was not confined to that famous religious center. Many *cenotes* in the region of Chan Kom are still invoked in the *ch'a-chaac*, or rain-bringing, ceremonies.[72] This may well be a survival from an older *cenote* cult. As we have already noted, stones were sometimes attached to the body to make it sink.

Sacrifices were followed by a feast in which the participants and probably the spectators of rank took part. The food offerings were eaten, *posole*, chocolate, and sometimes *balche* were drunk. A drunken orgy is specified in some cases, which took place in the church on at least one occasion. At Yaxcaba a memorial service took place about two months after the sacrifice in which food and drink were offered and consumed. This is said to have been an old custom. After eating, the priest elevated the idols and preached a sermon to those present.[73]

---

[70] Landa, 1928–1929, vol. 1, p. 202.

[71] Roys, 1933, appendix A.

[72] Redfield and Villa, 1934, p. 349.

We have already noted a disposition among the Maya to combine Christianity with their own native religion. They were ready to accept innovations, but they frequently fitted them into their own religious pattern. There were, indeed, a number of striking, if superficial, resemblances between the religious practices of the Spaniards and the Maya. Not only did the cross resemble an important Maya symbol, the Yax-cheel-cab, or First Tree of the World, but in both religions we find baptism, confession, ceremonial consumption of fermented liquor, fasting, and continence, to say nothing of such minor details as incense and altars. If certain missionaries were so impressed as to believe that Christian teaching had already reached the Maya at some very early time,[74] we may well imagine the effect these things produced on the native mind.

The present documents bring to light some of the earliest manifestations of this fusion of Christianity with paganism, which has continued down to the present time. We read of an offering of Spanish wine mixed with honey and water, honey being the chief ingredient of *balche*, the native sacred beverage. Also "thick candles" were burned at sacrifices, evidently the type used in Christian worship and not merely for illumination. Already in 1562 the Indians of the Sotuta area had for some time been crucifying their sacrificial victims. The latter were either tied with vines or nailed with wooden pegs to the cross and ceremonially named Jesus Christ. Once when two girls were sacrificed in this manner, the *ah-kin* who officiated said, "Let these girls die fixed to the cross as Jesus Christ died, who they say was Our Lord but we do not know he was." Another witness describing the same case reported that the *ah-kin* said, "Here you see the figure of Jesus Christ."[75] Sometimes the victim was removed from the cross before the heart was excised; at other times it was left on the cross when this was done, and the cross and body together were cast into a local *cenote*. On one occasion the corpse was carried on the cross to

---

[73] Cf. Landa, 1928–1929, vol. 2, p. 70.

[74] Saville, 1921, p. 214.

[75] Scholes and Adams, 1938, vol. 1, doc. 12.

Chichen Itza and thrown into the Sacred Cenote. According to the testimony of several witnesses, the crosses used for these sacrifices were kept in the house of the cacique of Sotuta.

Crucifixion was evidently associated with the worship of the rain gods and the *cenote* cult. One of the first missionaries, Francisco Hernandez, reported that the cross was adored as the god of water or rain,[76] referring no doubt to the Yax-cheel-cab.

In course of time the cross came to be regarded as a personality. In the "Títulos de Ebtun" the large cross near the town church is referred to as "Ca Yum Santa Cruz," or "Our Lord the Holy Cross." At Chan Kom Redfield reports that the cross is a *santo* and receives the same sort of worship as that accorded to the other *santos*.[77] As late as the middle of the nineteenth century we find the Santa Cruz Indians worshiping a plain wooden cross, two to three feet high. It was supposed to be endowed with the power of speech, and human sacrifice has been reported in connection with its cult.[78]

IV

It has often been stated that the inquiry of 1562 was the occasion of a wholesale destruction of Maya antiquities, especially priceless codices written in hieroglyphs. Consequently Landa's name has been anathema to many students of ancient American life. What are the facts as revealed by the documents?

Thousands of idols were destroyed. Most of these were small pottery objects and a few were of wood. So few references to stone idols are found in the record of the inquiry that we may doubt whether many important sculptured stone objects were lost at this time. The question of the codices is, of course, the important one. It has been reported, on the basis of early nineteenth century documents, that twenty-seven codices were burned during the investigation.[79] In view of the fact that the papers of 1562–1563 contain so many references to idols and that the friars cited the great numbers that were collected and destroyed as evidence of the guilt of the Indians, we should also expect to find numerous statements concerning books if many of the latter had been found. It is significant, therefore, that the record of the inquiry contains only one reference to "books." This is in the declaration of an Indian who gave evidence at Homun in September 1562. The document briefly states without comment that the witness admitted

---

[76] Saville, 1921, p. 209.

[77] Redfield and Villa, 1934, p. 110.

[78] Gann, 1918, pp. 41–42.

[79] Lopez de Cogolludo, 1867–1868, vol. 1, p. 604.

> . . . that he has had idols and books and masks, and that he brought and displayed four masks and some books.[80]

In his "Relación de las cosas de Yucatan," Landa states:

> These people used certain characters or letters, with which they wrote in their books about the antiquities and their sciences; . . . We found a great number of books in these letters, and since they contained nothing but superstitions and falsehoods of the devil we burned them, all of which they took most grievously, and which gave them great pain.[81]

We cannot be sure whether Landa refers here to the inquiry of 1562, or to a general practice adopted as part of the missionary program. It seems reasonable to believe, however, that if a large number of codices were destroyed in 1562, the contemporary record of the inquiry would contain explicit statements to that effect.

The idolatry proceedings of 1562 also had a direct relation to Landa's treatise on Maya antiquities. Landa had undoubtedly started collecting data for the "Relación" during the decade from 1550 to 1560, but apparently it was not until after he went to Spain to defend himself before the king and Council of the Indies that he actually whipped the materials into shape. At least, we may infer this from the fact that the only manuscript version of the "Relación" that has been preserved bears the date 1566, and from internal evidence in the work itself. It is also reasonable to suppose that one of his motives for writing it at that time was the fact that it would serve to illustrate those religious rites and idolatrous practices which he had so actively sought to extirpate.

In the same year that Landa was acquitted by the decision of the provincial of Castile, Philip II gave orders for the establishment of tribunals of the Holy Office of the Inquisition in Mexico City and Lima. Jurisdiction in all cases of heresy, blasphemy, bigamy, propositions contrary to Catholic doctrine, evil-sounding words, and other offenses against the Church was transferred from the bishops and other local prelates in America to these special courts. There had been numerous complaints against the activities of the bishops and prelates as inquisitors, but it may be doubted whether any of these had caused such a deep impression as the Landa case. Indeed, it seems reasonable to assume that the latter had a direct and immediate influence on the king's decision to establish separate tribunals for the trial of offenses against the Faith. Moreover, the laws defining the authority of these tribunals specifically exempted the Indians from their jurisdiction, because of their incapacity and lack of understanding concerning Christian doctrine and because many were not well instructed in the teachings of the Church.[82] This statement of policy vindicates the point of view held by Bishop Toral in 1562, and we may venture to suggest that the Landa episode was a contributing factor in the formulation of such a policy.

---

[80] Scholes and Adams, 1938, vol. 1, doc. 14.

[81] Landa, 1937, p. 82.

In retrospect the inquiry of 1562 is seen as an incident, albeit a spectacular one, of that conflict of cultures in Hispano-Indian America deriving from the Conquest, an illustration of the stresses and strains in native life produced by the introduction of European civilization. Not only in Yucatan, but also in other areas the forced acceptance of Christianity had an important influence on the course of colonial history. The Pueblo Revolt of 1680 in New Mexico was a movement to throw off Christianity and return to aboriginal customs and modes of life. The rebellion of the Tzendal Indians of Chiapas, inspired by the belief that an Indian girl had received messages from the Virgin and organized by another Indian who proclaimed "that St. Peter had taken him up to heaven and appointed him as his vicar on earth, with power to elect bishops and priests,"[83] illustrates another phase of the problem. Other examples of the interaction of Christian and native religions could be cited for other parts of Spanish America. For students of Central American antiquities, however, the investigation of 1562 has special interest because of the rôle played by Fray Diego de Landa, the founder of Maya research.

LITERATURE CITED

Archivo General de Indias, Seville. Tribute Roll for the Province of Yucatan, 1548–1549, *in* Audiencia de Guatemala, leg. 128.

——— Paxbolon-Maldonado papers relating to the *cacicazgo* of Acalan-Tixchel, *in* Audiencia de México, leg. 138.

---

[82] Solorzano, 1648, bk. 4, chap. 24.

[83] Bancroft, 1886, vol. 2, pp. 696–706.

BANCROFT, H. H., 1886. History of Central America. 3 vols. San Francisco.

BRINTON, DANIEL GARRISON, 1882. Maya Chronicles. *Library of Aboriginal American Literature*, No. 1. Philadelphia.

CARTAS DE INDIAS, 1877. Madrid.

DOCUMENTOS PARA LA HISTORIA DE YUCATAN, 1938. Vol. 2: La Iglesia en Yuca tan, 1560–1610. Merida.

GANN, THOMAS W. F., 1918. The Maya Indians of Southern Yucatan and Northern British Honduras. U. S. Bur. Amer. Ethnol., Bull. 64, Smithsonian Inst.

——— and J. ERIC THOMPSON, 1931. The History of the Maya from the Earliest Times to the Present. Chicago.

HERNAEZ, F. J., 1879. Colección de Bulas, Breves y Otros Documentos Relativos a la Iglesia de América y Filipinas. 2 vols. Brussels.

LANDA, DIEGO DE, 1928–1929. Relation des Choses de Yucatan. Spanish text with French translation and notes by Jean Genet. 2 vols. Paris.

——— 1937. Yucatan before and after the Conquest, with Other Related Documents, Maps, and Illustrations. Translated with notes by William Gates. Maya Society, Baltimore.

LOPEZ DE COGOLLUDO, DIEGO DE, 1867–1868. Historia de Yucatán. 3d ed., 2 vols. Merida.

PEREZ CODEX, *ca.* 1837. MS owned in Merida, Yucatan.

PEREZ, JUAN PIO, 1866–1877. Diccionario de la Lengua Maya. Merida.

REDFIELD, ROBERT, and ALFONSO VILLA R., 1933. Chan Kom, a Maya Village. Carnegie Inst. Wash. Pub. No. 448.

RELACIONES DE YUCATAN, 1898–1900. *Colección de Documentos Inéditos Relativos al Descubrimiento, Conquisto y Organización de las Antiguas Posesiones Españolas de Ultramar*, vols. 11 and 13. Madrid.

ROYS, RALPH L., 1931. The Ethno-Botany of the Maya. *Middle Amer. Research Ser.*, No. 2, Tulane Univ. of Louisiana, New Orleans.

——— 1933. The Book of Chilam Balam of Chumayel. Carnegie Inst. Wash. Pub. No. 438.

SAHAGUN, BERNARDINO DE, 1938. Historia General de las Cosas de Nueva España. 5 vols. Mexico, D. F.

SANCHEZ DE AGUILAR, Pedro, 1892. Informe contra Idolorum Cultores del Obispado de Yucatan. Anales Museo Nac. de México, vol. 6, pp. 13–122.

SAVILLE, MARSHALL H., 1921. Reports on the Maya Indians of Yucatan, *Indian Notes and Monographs*, Mus. Amer. Indian, Heye Foundation, vol. 7, pp. 137–226. New York.

SCHOLES, FRANCE V., and ELEANOR B. ADAMS, 1938. Don Diego Quijada, Alcalde Mayor de Yucatan, 1561–1565. 2 vols. *Biblioteca Histórica Mexicana de Obras Inéditas*, vols. 14 and 15. Mexico, D. F. (In press.)

SELER, EDUARD, 1902–1923. Gesammelte Abhandlungen zur Amerikanischen Sprach- und Altertumskunde. 5 vols. Berlin.

SOLORZANO Y PEREYRA, JUAN DE, 1648. Política Indiana. Madrid.

THOMPSON, J. ERIC, 1930. Ethnology of the Mayas of Southern and Central British Honduras. Field Mus. Nat. His. Pub. 274, *Anthropol. Ser.*, vol. 17, no. 2.

TIZIMIN, BOOK OF CHILAM BALAM OF. MS, original in Mexico, D. F. Reproduced by Maler and Gates.

PALÆONTOLOGY AND GEOLOGY

STUDIES RELATING TO EARLY MAN IN AMERICA

E. B. HOWARD

Academy of Natural Sciences of Philadelphia
Research Associate, The Carnegie Institution of Washington

It seems particularly appropriate at this time to consider briefly the progress made in studies relating to the field of Early Man—studies in which Dr. Merriam has been interested for many years. Much of this progress, both here and abroad, can be attributed to Dr. Merriam's cooperation and to the inspiration and encouragment which he has given to those whose efforts have been directed toward shedding new light on the problem as a whole. This is not to intimate that the investigation of a problem of such broad proportions and complexities as that relating to Man's early history has progressed steadily. There are still a number of fundamental questions that remain unanswered.

Of these questions that of Man's antiquity in the New World is one of the most debatable. At the same time it is one of the most interesting and provocative phases of the whole problem and continues to challenge the best thought of scientists in many different fields of work.

Cooperation is undoubtedly the keynote to the solution of the problem, if solution there be. In recent years there have been exhibited among individual investigators, and institutions as well, a fine spirit and willingness to cooperate. Foundations with funds for research have been generous in promoting this sort of research. The Carnegie Institution of Washington under Dr. Merriam's direction has been foremost in promoting this spirit of cooperative research among geologists, anthropologists, palæontologists, and archæologists, all of whom have been absorbed in one way or another in attempting to piece together the picture of Man's early history in the Western Hemisphere.

In the past year outstanding results have been achieved. Dr.

deTerra, assigned to the Academy of Natural Sciences of Philadelphia, and supported by funds from the Carnegie Institution, has been enabled to bring to practical completion the report on his work in India, which has been of great importance in throwing new light on Man's beginnings in Asia. With the additional cooperation of the American Philosophical Society and the Peabody Museum of Harvard University he has been able to pursue original lines of study of the Pleistocene geology of Burma, and through Dr. Movius, representing the Peabody Museum, to throw new light upon the archæology of Burma as well. The results of these investigations are extremely important and will add much to the knowledge of Early Man in Asia, especially since Teilhard de Chardin, who has been working for a number of years at the *Sinanthropus* site in China, was a member of the party in Burma. Further than that, at the conclusion of the work in Burma, and again through the aid of Dr. Merriam, deTerra and his party had the exceptional opportunity of visiting Dr. von Koenigswald in Java.

To mention one other outstanding piece of work of the past year, due directly to Dr. Merriam's interest and cooperation, it has been possible for Dr. von Koenigswald to pursue his studies on the problem of *Pithecanthropus erectus* in Java. Discoveries within the past year of additional material from the banks of the Solo River in Java have led Dr. von Koenigswald to assign this new material to *Pithecanthropus* and thus have led to a reconsideration of the whole problem of the "Java ape-man."

The opportunity thus afforded, of having these young men get together in the field at several of the outstanding sites where discoveries of the remains of primitive Man have been made, cannot be overestimated from the point of view of cooperative research, the results of which are already manifest.

Less startling, but no less important, are the studies being made in America along similar lines of cooperation. Every opportunity that has arisen has been embraced where it has appeared that interesting and important discoveries might result, and in recent years Dr. Merriam has offered help in innumerable ways to further such opportunities.

An example which will illustrate this occurred at the time of the last meeting of the International Geological Congress at Washington, D. C., in 1933. The excavations and investigations which were in progress at the gravel pit near Clovis, New Mexico, were producing interesting human and faunal associations, and Dr. Merriam thought that results could be advanced materially by holding a field conference at the site. Accordingly he invited Sir Arthur Smith Woodward and Lady Woodward, and Dr. V. Van Straelen, director of the Natural History Museum of Brussels, who had been attending the meetings in Washington, to go to Clovis, where Dr. Chester Stock, of the California Institute of Technology, and several others were invited to join the party, and with Dr. Merriam they were able to pass judgment on the Clovis site then and there. The following season Dr. Merriam sent Dr. Antevs to the same site for the purpose of studying the geology of it and thus arriving at the age of the associated Folsom artifacts and remains of extinct animals.

Much research work in Arizona, California, Nevada, Oregon, Nebraska, and other states, upon the general problem of Man's antiquity in America, has been carried on in the same way in recent years under direction of Dr. Merriam.

While there are many phases to this subject, one of the most fundamental and at the same time elusive questions to be answered has to do with the length of time Man has been in the New World and what his origin may be. There is no definite answer yet to this, but the problem is being attacked with renewed energy by present-day workers. Though we may never have a complete picture of early human history in America, it is well perhaps to examine the problem from time to time in the light of the more recent discoveries and to consolidate what gains may have been made toward a solution. During the past decade so much has been done that we wonder whether we have paused long enough to assimilate some of the many facts that have been appearing in such rapid succession during this period.

It may be of interest, therefore, to see how the time element of our problem now stands.

When the discoveries near Folsom, New Mexico, were announced somewhat over ten years ago, a great impetus was given to the study of Early Man in America. For many years previously investigators along this line had been so discouraged that they had turned away from attempts to solve the question of antiquity and were devoting time to studies of cultures that were relatively closer to the recent Indian. Startling discoveries had been announced at various times previous to the past decade and dating back to the work done by such of the earlier investigators as Abbott and others. Many of these discoveries purported to show evidence for Man's appearance in America at periods corresponding to the Palæoliths of Europe, and in South America, under the enthusiastic Ameghino, Man was even placed in the Tertiary.

It is not our purpose to review the reports of these various discoveries covering the first half of the last century. These have all been reported in numerous publications over a long period. What we are interested in doing is to take stock of the conclusions drawn from reports of discoveries made in more recent years. We realize that there is still much disagreement and speculation as to dating except as applied by dendrochronology. About all that can be done is to take the estimates of the age of these discoveries, as given by the best authorities on the subject, for the purpose of giving us a picture of what has been done and what remains to be done to establish a working American chronology.

The Southwest offers the best opportunity to start our stocktaking. Here we can work backward from dates that are well authenticated by tree-ring studies. Until recent years, after these studies by Dr. Douglass and his associates really got under way, chronology of the prehistoric Pueblo and Basket Maker people was thought to have extended much farther into the past than seems to have been the case, but as Roberts points out, cultural developments do not appear to have taken nearly so long as had been estimated. It had been thought that the Basket Maker, for example, had lived in the San Juan region as long ago as 1500–2000 B.C. In the latest estimates Kidder brings these forward by many centuries.

The Pueblo series as given by Roberts may be indicated as follows:

Pueblo V (Historic Period) extended from present to 1700
Pueblo IV 1700–1250
Pueblo III (Great Pueblo) 1250 to end of tenth century

Pueblo II	Developmental Pueblo	end of tenth century to latter part of ninth century
Pueblo I		latter part of ninth century to end of seventh century

According to Kidder's estimates, modified Basket Maker (Basket Maker III) was under way in the San Juan by the latter part of the fifth century, to last in that region till the beginning of the ninth century; Basket Maker (Basket Maker II) extending perhaps not more than a century or two before modified Basket Maker. Colonial Hohokam seems to equate chronologically with late modified Basket Maker, and according to Kidder's estimates the first pre-Colonial Hohokam culture runs back to about the beginning of the Christian era. We may note here, as a matter of interest, that there is evidence that the Pueblo series is punctuated with droughts of varying intensity beginning with A.D. 790 up to the great drought that lasted from 1276 to 1299.

So far, then, we have taken our chronology back to about A.D. 1 in the Southwest where in southern Arizona a pottery-making people, the Hohokam, had already appeared on the scene, while the Basket Maker of northern Arizona in a non-pottery-making, maize-growing stage appears at a somewhat later period, perhaps three or four centuries after the beginning of the Christian era, according to Kidder's interpretation. How much farther back is there evidence that Man existed in North America?

It may be well to note here some of the more recent sites where evidence points to earlier occupation of the country than A.D. 1. Among these may be mentioned, in New Mexico, Folsom, Conkling Cave, Burnet Cave, Clovis; in California, Lake Mohave, Pinto Basin, Lake County; in Arizona, Cochise; in Texas, Abilene; in Nebraska, Signal Butte, Scottsbluff Bison Quarry; in Colorado, Lindenmeier, Dent; in Minnesota, Pelican Lake; in Nevada, Gypsum Cave; and in Florida, Vero-Melbourne. There are other sites in Nebraska, Colorado, Texas, and Minnesota which can probably be correlated with those given, but which do not throw any additional light on the subject.

Kirk Bryan considers the valley-fill or flood-plain deposits at Folsom, New Mexico, containing extinct bison remains, as early recent or very late Pleistocene. This period is variously estimated by geologists at from 10,000 to 20,000 years ago. Conkling Cave, though more difficult to date, may be roughly the same age based on the faunal assemblage, according to Antevs and others. Burnet Cave supplied stratigraphic evidence in having a culture a few feet below the surface resembling Basket Maker, while below this a

Folsom culture with associated faunal remains places this latter at about the same age as the Folsom site. The Clovis site is dated by Antevs at 12,000–13,000 years.

The same author dates the Lake Mohave sites at 15,000 years, and Amsden places the Pinto Basin site at 10,000 years. The Lake County site has been recently reported upon by Harrington as having Folsom artifacts above an older culture as yet undated. The Cochise culture Antevs correlates roughly with the Lake Mohave sites, which he dates at about 15,000 years.

The earlier sites at Abilene, Texas, have been studied by Antevs and also by Leighton. The latter tentatively concludes that primitive Man lived in the Abilene region just before, and during, the last or Wisconsin glacial stage.

At Signal Butte, Nebraska, Strong finds three levels which he indicates, subject to later correction, as follows. Level I falls either in the dry cool period about 8000 years ago or in the humid cool period about 10,000 years ago. Level II falls in the humid period of 5000 years ago, and Level III into the humid period of the last 1000 years. This last level contained prehistoric pottery. Thus the lower level (I) might be placed a little nearer the present than some of the Folsom sites mentioned above. The Scottsbluff Bison Quarry showed association of extinct bison and artifacts similar to that at other sites already mentioned, and probably dates from about the same time, namely somewhere around 10,000–15,000 years ago. Other Yuma sites have been discovered recently by parties from the University of Nebraska under C. B. Schultz's field direction, and a more definite dating of these new sites will be looked for in the near future.

The Lindenmeier site near Fort Collins, Colorado, has been extensively studied by Roberts, who is inclined to date the site at roughly the same age as that of Clovis, although more recent geological studies indicate that the deposits may be considerably earlier. At Dent, Colorado, where more or less the same pattern of association is found as at other sites mentioned, it would seem to be safe to give the same date, namely the beginning of postglacial times.

The discovery near Pelican Lake, Minnesota, has caused considerable controversy. If the majority opinion prevails the human remains date back to a period when the varved clays were still forming at the end of Pleistocene times. If one holds with the minority viewpoint, then the deposits in which the human remains were found were disturbed and no date can be placed upon them.

Dr. Harrington estimates the earlier occupation and association with the ground sloth in Gypsum Cave at 10,000–13,000 years.

The Vero-Melbourne sites present difficulties in dating due to strenuous minority objections to the validity of the associations. The faunal pattern, however, contains so many of the same forms as that at many of the sites already mentioned that it does not seem logical to exclude it from the list. On this basis it probably correlates with other late Pleistocene or early postglacial sites.

Thus, so far as North America is concerned, we have a number of recent discoveries, well authenticated, that date from somewhere around the closing stages of the Wisconsin glacial period or the early part of postglacial times. This might be from 10,000 to 15,000 years ago. With the exception of Signal Butte Level II we have practically no evidence to show what comes in between say 10,000 years ago and A.D. 1 as applied to North America. In Central America, according to Kidder, a settled, agricultural existence of the Mayas began about 400 B.C. and the beginnings of American civilization date from about 1000 B.C., prior to which a hunting and food-gathering stage extended beyond. How can we fill up this enormous gap? And when we have completed the earlier part of our North American chronology who will say where our Mound-builders belong and the so-called Red Paint peoples and the many other peoples whose records are left for us to decipher?

It looks like a discouraging task, but in view of the advance in recent years toward a clearer picture of our problem it cannot be said to be hopeless.

ANCIENT FORESTS OF OREGON: A STUDY OF EARTH HISTORY IN WESTERN AMERICA

RALPH W. CHANEY
University of California
Research Associate, The Carnegie Institution of Washington

Introduction

The student of earth history is constantly searching for new evidence regarding the development of the planet on which he lives. He is able to read only a small fraction of the complete story from the formally written record, since man arrived upon the earth long ages after its beginning. Continents have been lowered so that seas advanced widely across them; mountain ranges have risen up along coastal lowlands; in a single region the climate may have ranged from tropical to boreal, from arid to humid. Such events and trends are recorded by layers of rock and the fossils contained in them. The book of the earth, as read by the geologist and the palæontologist, has for its pages sandstone layers and lava flows, for its illustrations the shells and bones of animals, the leaves and stems of plants.

Because land plants have always lived in intimate contact with the soil, reflecting the temperature and moisture of their day and the topography of their terrain, the study of palæobotany throws much light upon the nature of the earth in the past. In western America vulcanism has provided a setting especially favorable for the preservation of plant fossils. Volcanoes mantled wide areas with ash and tuff during the Tertiary period,[1] burying the forests of

---

[1] The latest era of earth history, during which the principal development of modern plants took place, is classified by geologists as follows:

Cenozoic era	Quaternary period	Recent epoch
		Pleistocene epoch
	Tertiary period	Pliocene epoch
		Miocene epoch
		Oligocene epoch
		Eocene epoch

the lowlands and filling lakes and valleys. In these deposits leaves, fruits, and occasional flowers have been preserved as impressions; hot volcanic waters have penetrated trunks of buried trees, leaving behind the silica which they bore in solution and transforming the wood to stone; sheets of lava poured out and hardened to form a protective capping beneath which much of this record of past vegetation has remained down to the present. Differing but little from trees and shrubs of today, these fossils may be identified and interpreted by comparing them with modern vegetation whose topographic and climatic requirements can be directly observed.

A striking modern example of the control exerted by topography and climate may be seen in the Cascade Range of Oregon. From a dense and humid forest of giant hemlocks and firs on the west slopes, there is an abrupt change to an open pine woodland east of the summit along the Mackenzie Highway; descending into the great interior plateau, pines give way to scattered junipers and desert scrub. The student of vegetation, whether his interest centers upon living or fossil plants, finds in such a situation the key to many of the problems involved in plant distribution. Unlike the botanist, who must confine his observations largely to the spatial aspects of plant distribution, the palæobotanist may also consider vegetation changes in time. He has at his disposal nearly one hundred million years of earth history during which plants of modern types have been in existence. In the region now occupied by the Cascade Range in Oregon, the fossil record indicates a constant change both of composition and of distribution, a response to a slowly changing earth. From a subtropical forest which extended over broad lowlands at the beginning of the Cenozoic era, there has developed the diversified vegetation which characterizes this now temperate region. It is with such changes, both in time and in space, that my studies have been concerned during the past two decades, largely under the auspices of the Carnegie Institution of Washington, and with the direction and close cooperation of its president, John Campbell Merriam.

In this attempt to trace the history of Oregon through a study of its past vegetation, reference has been made to several other lines of investigation. The margins of ancient seas have been traced by Weaver (1; see "Literature Cited" at end of paper), Packard, Clark, and Schenck (2) from the occurrence of marine animals left in sediments along Tertiary shores. Advice in reconstructing the habitats of land plants has been sought from Jepson, Mason, Merrill, and Standley; familiar with the distribution of modern vegetation, these botanists have been able to suggest the environments occupied by the ancestors of living trees. A survey of the origin, distribution, and structure of rocks gives further information regarding the valleys and mountains of the past, as studied by Williams, Callaghan (3), Hodge, Buwalda, and other geologists. Merriam, Furlong, Sinclair, and Stock have described the mammals which lived in Tertiary forests, or ranged along their borders. The various components of the picture of Oregon's past have been drawn from many sources.

The ecological studies carried on by Clements (4) in western America have constantly served as a guide both for method and for interpretation. In the preparation of palæogeographic maps, the principles employed by Schuchert (5) have been carefully followed, or adapted to the expression of terrestrial data. Finally, the point of view expressed by Merriam throughout forty-four years of teaching and writing has provided the basis of my approach to the problems of the earth and its life. Briefly stated, this viewpoint recognizes fossil animals as creatures which have lived, moved, eaten in the days of the past, creatures whose structure and migrations are a chronicle of conditions before the days of man. It recognizes fossil plants as a covering of the earth which has responded to moisture, temperature, and slope as do plants of today. In his essay "The Story of a Leaf" (6) Merriam has written of a fossil from the Miocene rocks of Oregon: "Through the assured reality of this leaf, as it was then and had been so long ago, a view was opened across the eons connecting present and past. In that sweep of vision it was not possible to avoid the panorama of history just as it had occurred."

TERTIARY FLORAS OF EASTERN OREGON

The presence of a subtropical forest in eastern Oregon during the Eocene epoch is indicated by abundant fossil leaves and fruits which have been collected from the Clarno formation (see table 1). The environment which this forest occupied was so different from any now to be found on the landward side of the Cascades, or any-

TABLE 1. TERTIARY SECTIONS IN THE CASCADE REGION

Geologic age	John Day Basin	Ashland region	West central Oregon	Eastern Washington
Middle to Lower Pliocene..	Rattlesnake and Deschutes formations, the latter with a cool-temperate flora	High Cascade lavas*	High Cascade lavas	Rhyolitic lavas
Upper to Middle Miocene..	Mascall formation with a temperate flora containing some warm-temperate elements	Western Cascade lavas, including volcanic ash with a flora resembling the Mascall	Western Cascade lavas	Ellensburg formation with a flora resembling the Mascall
Middle Miocene..........	Columbia River lava	Western Cascade lavas	Western Cascade lavas	Lavas (Keechelus, Yakima) including volcanic sediments, Latah formation with a flora of mixed climatic types
Lower Miocene to Upper Oligocene..............	John Day formation containing the temperate Bridge Creek flora	Western Cascade lavas† including volcanic sediments containing the temperate Kincaid Ranch flora	Western Cascade lavas, including volcanic sediments containing the Rujada and Cascadia floras of mixed climatic types	
Lower Oligocene to Upper Eocene................			Fisher formation containing the subtropical Goshen flora	
Upper Eocene...........	Clarno formation with a subtropical flora	Fresh-water and volcanic sediments corresponding to the Comstock formation with a subtropical flora	Comstock formation with a subtropical flora	Roslyn and Manatash formations with a subtropical flora

* Post-Tertiary flows are also included in the High Cascade lavas.
† Certain of the Western Cascade lavas may be as old as Upper Eocene.

where else in temperate America, as to indicate major changes in climate and topography during the ages since the Eocene. In the period preceding, known as the Cretaceous, an embayment extended inland from the Pacific to eastern Oregon. Its presence across the John Day Basin is shown by the occurrence of fossil shelled animals, including the now extinct ammonites, in dark gray shales laid down as mud in the Cretaceous sea. It is not known whether this embayment continued eastward to join the inland sea which covered the Rocky Mountain area of today, but there are many suggestions that there was at least a lowland lying between the two areas. When the Rocky Mountains were folded up in late Cretaceous time, the sea withdrew from the interior of North America. At the same time the Oregon embayment was limited to the western part of the state. The dawn of the Cenozoic era found Oregon a broad plain across which the Cascade range had yet to be uplifted.

Such a statement can be made with a high degree of confidence even though no explorer was to map the Oregon country until tens of millions of years later, and although no photograph of an Eocene landscape can be used for verification. It is based upon a comparison of the forests which occupied eastern Oregon with those which lived to the west during Eocene time. Modern forests differ greatly from east to west, owing to the topographic and climatic influence of the Cascade Range, but in Eocene time the vegetation across the state was essentially uniform. The avocado (*Persea*), cinnamon (*Cinnamomum*), fig (*Ficus*), and persimmon (*Diospyros*) found in the Goshen flora near Eugene (7) or in the Comstock flora to the south (8) are identical with those which lived in the John Day Basin. Of more than forty species of trees, shrubs, and vines which make up the Clarno flora, there are only four or five which have not been found in older Tertiary rocks along the Pacific slope from Washington to California. The conclusion is inescapable that the mild and humid climate of the coastal borders extended inland to the John Day Basin of eastern Oregon, over the area now occupied by the Cascade Range. For if a mountain barrier had divided Oregon during the Eocene, the fossil floras would have been as different as they are across the state today.

Since the writer first came to realize the subtropical character of these floras over ten years ago, he has made two trips to Central and South America and another to Mexico, in search of the settings in which their modern equivalents have survived. In the countries to the south he has found living forests which represent, in somewhat modified form, the Eocene vegetation of western America. Growing on the lower slopes of the Andes in Venezuela, on the savannas of Panama, and in the mountains of Costa Rica and Guatemala there are trees with leaves, fruits, and stems like those of the Clarno and other Eocene floras. Over a span of some sixty million years, they have migrated southward in response to a changing earth. Their survival in the tropics, where the climate is humid and mild, gives us a picture of the Oregon landscape before the Cascades were built, a picture imprinted upon Eocene rocks which has lasted through the ages.

A map has been constructed to show the appearance of Oregon and Washington during the later days of the Eocene. The shore line is based upon the occurrence of marine fossils[2] and of coastal forests; valleys and rivers are indicated by the presence of fossil plants like those which now live on lowlands in moist regions. Localities where plant fossils have been found are indicated by numbers, as listed below figure 1. The nature of the vegetation on the uplands above the valleys is incompletely known; conditions for the preservation of fossils are seldom favorable at high altitudes where erosion is active and where there are few opportunities for leaves and stems to be buried in sediments. Fortunately there is a layer of leaf-bearing ash in the Clarno formation near the town of Fossil which appears to represent the vegetation of the hills. A small collection includes leaves of a few lowland trees mingled with the remains of typically temperate species such as redwood, alder, tan oak, and elm. A similar forest is known to have lived in Alaska during the Eocene. Just as the aspens and spruces common today on Arctic lowlands range southward into the mountains of middle latitudes, so the northern forest of the Eocene may be supposed to have been extended as far south as Oregon in the uplands. The presence of hills or mountains is therefore suggested by the mixture of this temperate flora with the more tropical species of the valley below.

Evidence will be presented below to show that certain elements of the lowland subtropical forest lingered on in western Oregon until a later epoch. But in the John Day Basin the passage of time brought their complete disappearance; forests of the early Miocene were of a temperate type, like those of today in the valleys of Michigan and Ohio, and in the Redwood Belt of California; they were essentially like those which had lived in the uplands during the Eocene, as judged by the small Clarno collection just mentioned. It must have been a dramatic moment in the life of Thomas Condon,

---

[2] This is adapted from a map recently published by Weaver, as are the shore lines on other maps in the present paper.

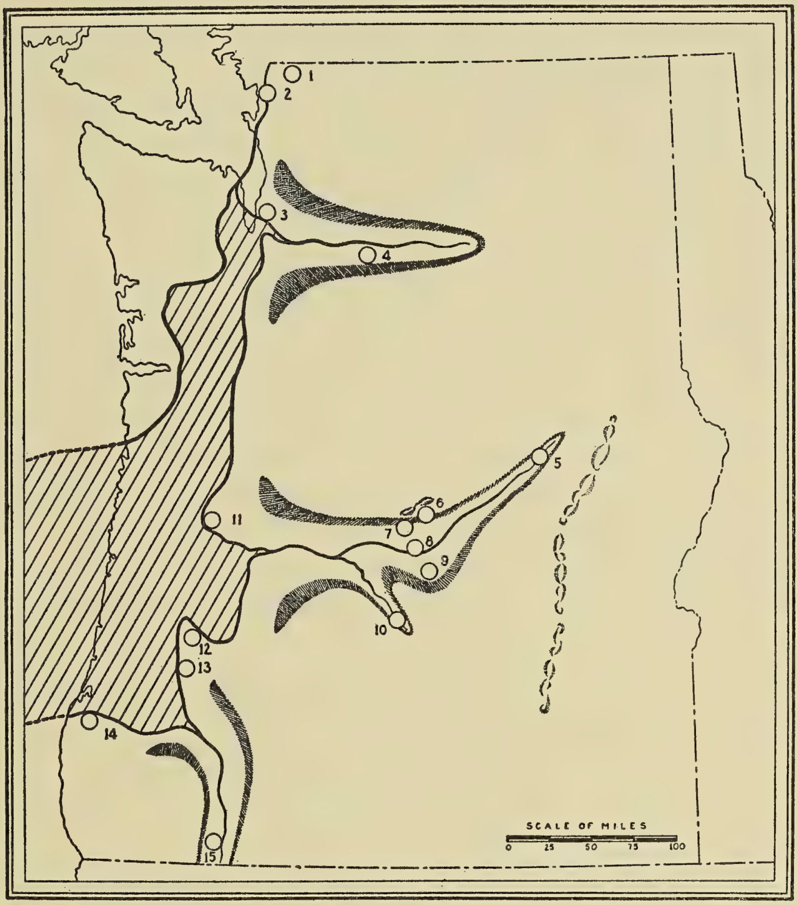


FIG. 1. Palæogeographic map of Oregon and Washington during later Eocene time. Seas are represented by area covered by diagonal lines. Open circles indicate plant localities numbered as follows: (1) Nooksack Canyon; (2) Chuckanut Drive; (3) Steel's Crossing; (4) Roslyn; (5) Pilot Rock; (6) Fossil; (7) Clarno; (8) Cherry Creek; (9) West Branch Creek; (10) Riverside Ranch; (11) Crabtree Creek; (12) Goshen; (13) Comstock; (14) Arago; (15) Ashland.

the Oregon pioneer, when he first discovered the fossil leaves and cones of redwood (*Sequoia langsdorfii*) in the John Day Basin (9). Or perhaps he did not realize how incongruous a setting was the hillside where he collected these fossil plants in 1865 and 1866. Traveling with a company of United States cavalry, for the young state of Oregon was still occupied by hostile Indians, he found prints of redwood and other leaves preserved in reddish shale on a low hill above Bridge Creek. These were sent east to John Strong Newberry, of Columbia University, who recognized their importance and later wrote a monumental work on the Condon collection and other specimens.

During the years which have followed, many geologists and palæontologists have visited this region, today a prosperous ranching country. The first adequate study of its stratigraphy was carried on by Merriam during the field seasons of 1899 and 1900 (10). At the same time the fossil plants were studied by Knowlton (11). The beds containing the Bridge Creek flora are now referred to the lower member of the John Day formation, rocks made up of volcanic ash and sediments which were deposited in eastern Oregon at the dawn of the Miocene epoch, some 40 million years ago.

On the many occasions during the past eighteen years when the writer has walked up the valley of Bridge Creek to the fossil locality, he has never failed to sense the contrast between the sparse sagebrush cover of today and the redwood forest which once lived upon this slope. Nowhere are the results of a changing earth more apparent. Here is a region now semiarid and unforested. When redwoods stood upon this slope, scattering their leaves and cones to be buried in the valley, the climate must have been mild and moist, with only slight seasonal change. For the redwood of today grows only in coastal California and Oregon, where the tempering effects of the Pacific and its rain-bearing winds provide a suitable environment. The mountain range which now deprives eastern Oregon of such marine benefits could not have risen to the west before the Miocene, or redwoods would have been absent from the Bridge Creek flora as they are today on the arid hills of the John Day Basin. The factors which brought about this change are the factors which control the development and distribution of life in the world as we know it.

The fossil flora from Bridge Creek has been discussed so fully that there is little need for more than a summary of its characteristics at this time (12, 13, 14). Predominant is the redwood forest element, with such abundant genera as alder (*Alnus*),[3] dogwood (*Cornus*), hazel (*Corylus*), maple (*Acer*), Oregon grape (*Odostemon*), pepperwood (*Umbellularia*), redwood (*Sequoia*), and tan oak (*Lithocarpus*). There are also included trees and shrubs which now occupy the borders of this forest in California, such as ash (*Fraxinus*), cherry (*Prunus*), hackberry (*Celtis*), live oak (*Quercus*), madrono (*Arbutus*), pine (*Pinus*), plane tree (*Platanus*), rose (*Rosa*), willow (*Salix*). If rhododendron were included, there would be missing not one of the common genera of the modern redwood forest. Walking in this Miocene valley forty million years ago, we should have seen several trees which no longer live in western America, and others which are today confined to Asia. Basswood (*Tilia*), beech (*Fagus*), chestnut (*Castanea*), elm (*Ulmus*), hickory (*Carya*), and hornbeam (*Carpinus*) are now limited on this continent to the region east of the Rockies; most of them are distributed in eastern Asia as well. These trees, characterized by broad, deciduous leaves, are largely restricted to regions of summer rainfall. The most distinctive tree now confined to Asia was the katsura (*Cercidiphyllum*), which occurs today only in forests and temple grounds of Japan and China.

This temperate assemblage, grouping trees now found in the coastal forest of California with those living in eastern North America and eastern Asia, is no longer in existence in its Miocene form. Climatic and topographic changes in eastern Oregon have shifted most of it to other parts of the Northern Hemisphere; the redwood element has survived in a coastal region still characterized by mild, moist climate; the deciduous genera for the most part occupy eastern sides of the northern continents where summer rainfall is available. There remain in the vicinity of the John Day Basin only a few representatives of the Bridge Creek flora, including species of alder, cherry, dogwood, hackberry, maple, and willow. These hardy trees and shrubs are largely confined to stream borders in the Blue Mountains.

Withdrawal of the humid forest from the John Day Basin appears to have been gradual. Redwoods were still represented, but in greatly reduced numbers, in the Mascall flora of later Miocene time (15, 16, 17). Living with them were many black and live oaks, box elders (*Acer*), madroños, plane trees, and poplars (*Populus*). The modern descendants of most of these trees may be found today on the borders of the redwood forest near its eastern and southern margins, where moisture is limited. They took the place of trees of the earlier forest which needed a larger amount of rainfall than the environment of eastern Oregon afforded during the closing days of the Miocene. Already volcanoes were building a line of peaks to the west, and the interception of rain-bearing winds from the Pacific was under way.

---

[3] Since references are given to the papers in which these plants are described in detail, it is not considered necessary to include here their specific names.

The actual date of the building of the Cascade Range can be readily determined by the disappearance of the redwood and other moisture-requiring trees at the end of the Miocene. During the Pliocene epoch which follows, there is no record of any of them in eastern Oregon. The known vegetation (18) is limited to aspen and cottonwood (*Populus*), box elder, cherry, and willow, all of which are still surviving in favorable situations along streams in the area east of the Cascades. Corroborative evidence is furnished by the geologist, whose study of the sequence and structure of the beds folded up into the Cascades indicates that active mountain-building began at the close of the Miocene. From that time on, only plants able to live under extremes of temperature and precipitation could occupy the area to the east. The moderating effects of the Pacific Ocean were cut off by rising peaks which have made the interior of western America a land of steppe and desert down to our day.

Miocene Floras of Western Oregon

In contrast with the abundant record of fossil plants in eastern Oregon, Miocene floras on the west side of the Cascades are comparatively little known. Their small representation here appears to be due largely to the heavy cover of modern forests, which in their luxuriance have obscured the record of past vegetation. Construction of highways and trails during the past several decades has exposed an increasing number of plant-bearing strata, and the development of this forest in western Oregon can now be summarized.

In only one area west of the Cascades is a sequence shown from the older and warmer floras of the Eocene to the temperate vegetation of the Miocene. This is in the vicinity of Ashland, in the southern part of Oregon, where plant-bearing Eocene sediments are overlain by a thick series known as the Western Cascade lavas, which contain lenses of ash and volcanic shale. The lower of these on the Kincaid Ranch contains a flora which closely resembles the Bridge Creek assemblage. Redwood, alder, and tan oak are mingled with chestnut, elm, and hickory. New plants include an abundant poison oak (*Rhus*) and a buckthorn (*Rhamnus*) as yet undescribed. Modern equivalents of both these live today in the redwood forest or on its borders. A third new species resembles the sugar pine found at higher altitudes in Oregon and California today. Rarity of this pine in the fossil record suggests that its Miocene habitat was also remote from the valley bottoms, so that only an occasional bundle of needles was carried down to the areas of deposition. With these three exceptions, all of the twenty species of plants now known from the Kincaid Ranch flora are found also in the Bridge Creek flora. The occurrence of both floras in volcanic shales which have a similar position in the sequence of Tertiary rocks makes it possible to conclude that they are of essentially the same age, Lower Miocene. The extension of a redwood forest eastward from Ashland to the John Day Basin is consistent with our conclusion, as expressed above, that the Cascade Range did not serve as a barrier until a later date in the history of Oregon.

Northward along the west slope of the Cascades, plant fossils have recently been found at two new localities, near Rujada and Cascadia, through the cooperation of Dr. W. D. Smith of the University of Oregon. They occur in volcanic shales associated with lava flows as in the case of the plant-bearing deposits from Kincaid Ranch and Bridge Creek. Their reference to the Miocene is further substantiated by the discovery of such members of the redwood element as alder, Oregon grape, redwood, and tan oak, and by such deciduous trees as basswood, chestnut, elm, and hickory. But together with these typically temperate species of the Miocene, there occur leaves of avocado, catalpa (*Catalpa*), lancewood (*Ocotea*), and palmetto (*Sabalites*), whose living equivalents are found in warm-temperate to tropical regions. Nowhere in the world today is the redwood found growing in association with such warm-temperate trees. It is clear that the Miocene forest of west central

Oregon occupied an environment unlike that at Ashland and the John Day Basin during this epoch of the past, and likewise different from that of the modern redwood forest. Here is a problem whose solution may supply critical details regarding Oregon in the past, its coast line, its valleys and mountains, its climate. From the fossil evidence now available, the following suggestions may be made.

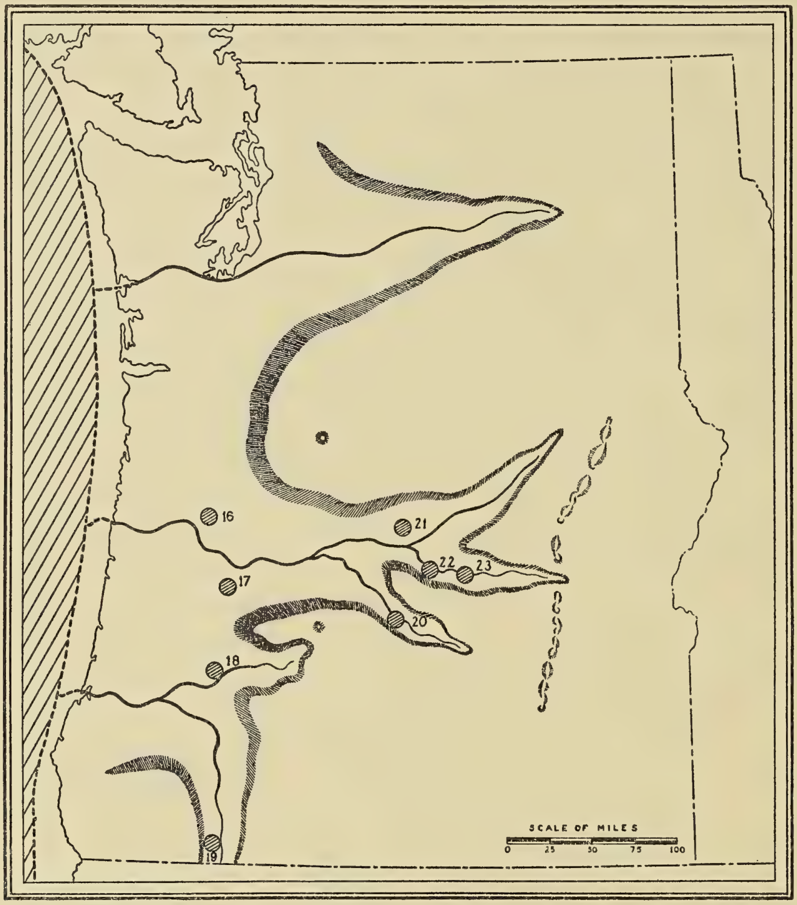


Fig. 2. Palæogeographic map of Oregon and Washington during early Miocene time. Seas are represented by area covered by diagonal lines. Circles with diagonal lines indicate plant localities numbered as follows: (16) Franklin Butte; (17) Cascadia; (18) Rujada; (19) Kincaid Ranch; (20) Gray's Ranch; (21) Clarno; (22) Bridge Creek; (23) Cant's Ranch.

On lowlands bordering a shore which appears to have been shifted westward beyond the present coast line, as shown in figure 2, there survived into the Miocene a number of trees which were members of the earlier subtropical forest. A similar survival may be noted today along the South Atlantic and Gulf Coasts, where representatives of the same genera (*Catalpa*, *Ocotea*, *Persea*, *Sabal*) have continued to live down to the present in the mild climate and sheltering topography of the ocean border. The Miocene equivalents of these trees lingered on in the coastal forest of Oregon after they had disappeared from more exposed habitats at higher elevations and farther inland.

On hills bordering the coastal lowlands a typically temperate forest lived under conditions less mild and humid. The present-day forest on the uplands near the coast in the southeastern United States is made up largely of pines. There is evidence that the upland forest adjacent to the Oregon coast was likewise made up of coniferous trees during the Miocene. A few twigs of redwood have been found in the shale at Rujada, twigs with scaly needles like those growing near the tops of modern redwoods; twigs from the treetops would be the most likely ones to be blown occasionally across the area intervening between the uplands and the lowlands, and to find their way into the stream deposits of the valleys. It seems probable also that with the advancing trend toward a colder climate, forests which lived in the interior were moving down the valleys toward the sea. The mixture of leaves of catalpa with those of tan oak and redwood is interpreted to represent changing conditions along the coastal plain, the gradual incoming of temperate trees to an environment which had sheltered the subtropical forest of Eocene. Farther up the valleys, where the Kincaid Ranch and Bridge Creek floras occur, the subtropical element had completely disappeared and the forest was strictly temperate in character. A similar situation may be seen today on the coast of Mexico and Central America, where the tropical forest of the lowlands gives way to typically temperate elements in valleys at higher elevations inland.

Although no redwoods live today on the border of warm and humid lowlands, it is of interest to note that an Asiatic relative, the Japanese cedar (*Cryptomeria*), may occur under corresponding conditions in the uplands of Japan. This tree is most common on slopes well above the warm-temperate forest of lower elevations, a forest containing trees which correspond closely to those of the warmer element of Miocene floras from the Oregon coastal lowlands.

However, it may range down into this lowland forest, and its twigs may become buried with the leaves of palms and laurels, as were those of the redwood in the Rujada flora. By frequently turning his attention to modern forests in which the relationships of living trees may be directly observed, the palæobotanist reaches a fuller understanding of forests whose settings are dimmed by antiquity.

In later Miocene time,[4] the return of the sea to a position east of the present coast is recorded by the occurrence of sediments containing marine fossils. As shown by figure 3, these form a narrow band along the Oregon coast, which extends inland more widely in central Washington. The return of a warmer climate is suggested by shellfish whose fossil remains have been found in these Middle Miocene rocks. No fossil plants have as yet been found along the coastal plain, but in the valleys extending eastward they are abundantly present. As previously mentioned, the Mascall flora of the John Day Basin contains the last redwoods known in the eastern part of the state; they are interpreted to indicate that the Cascades were not yet in existence, although there is evidence that volcanoes were already rising at isolated points along the axis of the present range. Oaks and madroños are numerous in the record, suggesting the proximity of hills to the sites of deposition. An element of the Mascall flora not previously discussed includes such genera as bald cypress (*Taxodium*), candelero (*Oreopanax*), mahogany (*Cedrela*), persimmon (*Diospyros*), and sweet gum (*Liquidambar*). Modern equivalents of these plants are found in regions far to the south of Oregon, where the climate is relatively mild. Their migration up the valleys into the interior of Oregon in later Miocene time fully confirms the evidence of a return to a warmer climate, as suggested by the marine fauna.

Even more suggestive of a warm-temperate climate is the Latah flora of eastern Washington. Somewhat younger than the coastal floras from Rujada and Cascadia, it contains many of the same species, including catalpa, avocado, and other warm-temperate trees, together with a temperate element of redwoods, tan oaks, and alders. An embayment of the Pacific in central Washington, nearly opposite the Latah occurrence, may well have extended landward to form a broad valley during Middle Miocene time. As shown in figure 3, this valley occupies the same general position as one drawn on the basis of Eocene floras (figure 1). Coastal elements may at the present time extend far inland in similar valleys, as in the case of the bald cypress, which ranges up the Mississippi Valley as far as southern Illinois. Separated from the Miocene ocean by a distance of nearly 300 miles, the Latah flora corroborates the evidence of the somewhat younger Mascall flora that the return of milder climate to western America brought about a landward migration of the coastal forest and its mingling with the temperate forest

---

[4] This epoch lasted for several million years according to estimates of geologists and palæontologists.

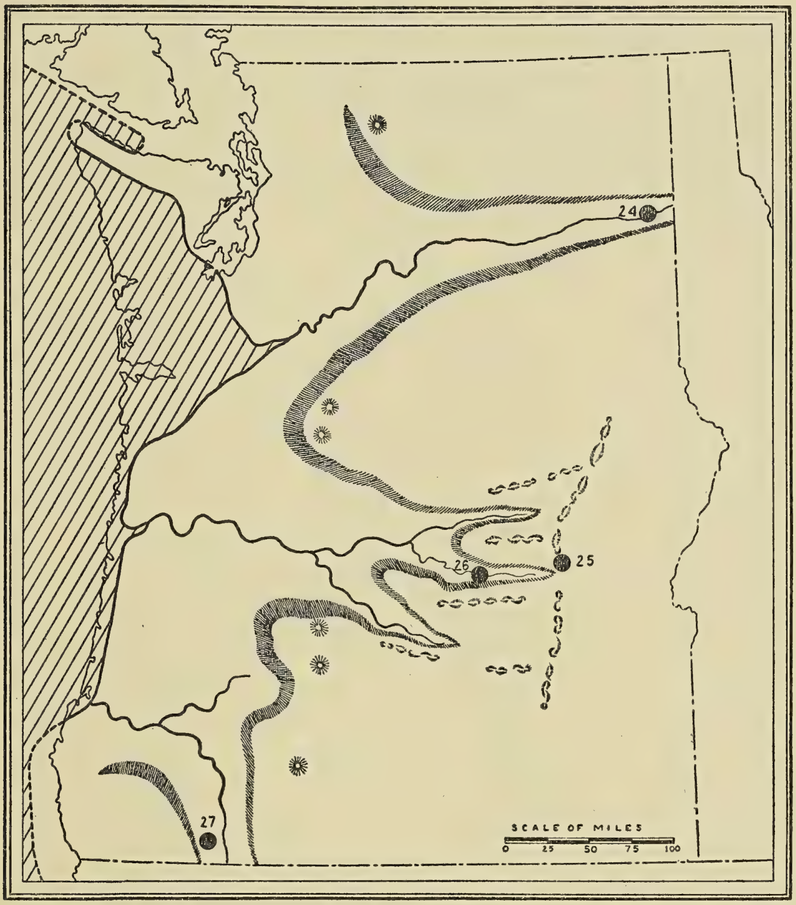


Fig. 3. Palæogeographic map of Oregon and Washington during middle to late Miocene time. Seas are represented by area covered by diagonal lines. Black circles indicate plant localities numbered as follows: (24) Latah; (25) Blue Mountains; (26) Mascall; (27) Grizzly Peak.

of the interior. Whether the redwood forests of Missoula, Yellowstone Park, and Florissant, recorded in beds of only slightly greater age, are to be interpreted as having occupied eastward extensions of these valleys of Washington and Oregon is a question which may be raised here without any attempt to marshall the evidence required for an answer. The Miocene was the last epoch in which warm-temperate coastal trees ranged inland to mingle with redwoods in the forests of eastern Oregon and Washington. Already major volcanic activity, and the mountain making associated with

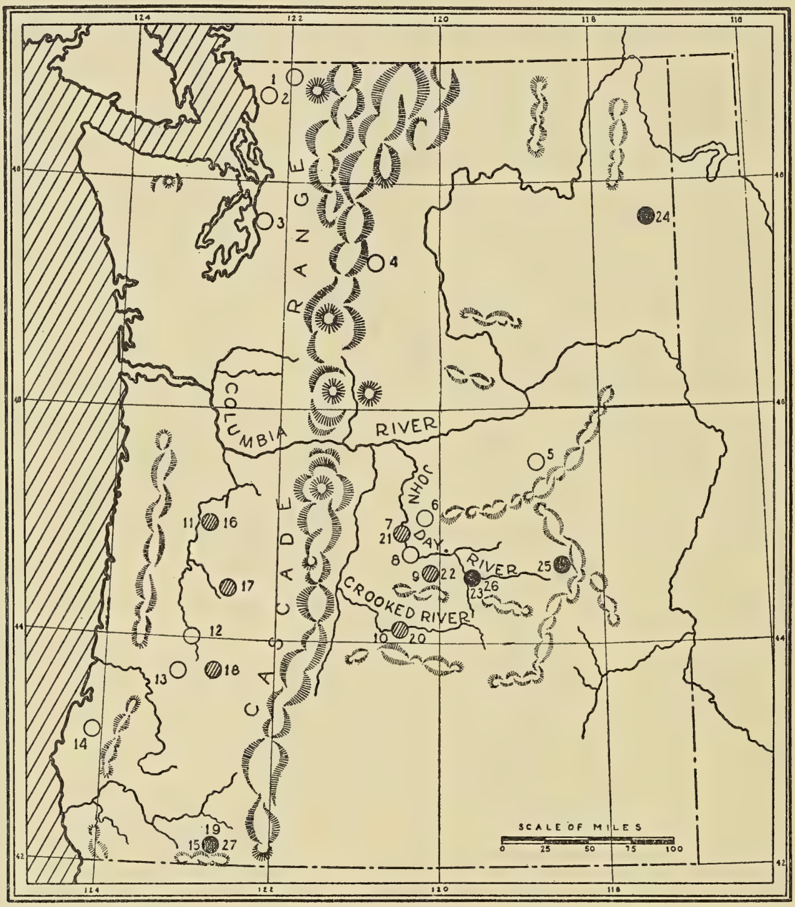

124
122
120
118
116
48
46
44
42
CASCADE RANGE
COLUMBIA RIVER
JOHN DAY RIVER
CROOKED RIVER
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
SCALE OF MILES
0
25
50
75
100

Fig. 4. Map of Oregon and Washington today. Plant localities are indicated by the legends of preceding maps. Certain localities have plants representing beds of two or more ages.

it, were producing profound changes in topography and climate. The western shore of our continent was assuming the form with which we are familiar at the present time.

Conclusion

The subsequent history of Oregon and its vegetation may be briefly stated. A line of volcanoes poured out lava flows which piled up and were folded to form the Cascade Range. No longer was there ready interchange between the west and east along broad valleys. In Oregon the drainage was turned northward, in Washington southward to the only large opening which has been maintained across the axis of the Cascades, the gorge of the Columbia River. Heavy rainfall along the coast has continued, but lowering temperature has shifted the catalpa and avocado southward. Douglas firs, oaks, and maples from the Miocene uplands now live along the coast, with redwoods surviving only from southwestern Oregon to central California. East of the Cascades rainfall has been limited, and the tempering effects of the Pacific are no longer felt. Sage and juniper have taken over this interior plateau, and only leaf impressions and logs of petrified wood remain to tell the story of the redwood forest which spread across the John Day Basin during the Miocene, and to record the presence of lancewoods and palmettos in the valleys of the Eocene.

Thus the sequence of events in earth history may be followed. Plant remains buried in the rocks provide evidence of the cutting of valleys, the building of mountains, the advance and retreat of the sea along ancient shores. Development, migration, and restriction of forests in western North America have been in response to earth changes. In these forests and upon their marginal grasslands, there lived the races of mammals which have so fully occupied our planet since the dawn of the Tertiary period. Man himself had his origin on the borders of forests migrating under the compulsion of changing environments. Constantly altered by geologic processes which build and destroy, the earth has ever presented a shifting scene to the plants and animals which have lived upon it. With the knowledge of the past furnished by ancient forests and their migrations, we may view with increased understanding the pages of history written as we live, and the chapters unfolding into the future.

LITERATURE CITED

1. Weaver, C. E., Tertiary Stratigraphy of Western Washington and Northwestern Oregon, *Univ. Wash. Pub. Geol.*, vol. 4, 1937.
2. Schenck, H. E., Marine Oligocene of Oregon, *Univ. Calif. Pub., Bull. Dept. Geol.*, vol. 16, no. 12, 1927.
3. Callaghan, E., Some Features of the Volcanic Sequence in the Cascade Range in Oregon, *Trans. Amer. Geophysical Union* (Nat. Research Council), April 1933.
4. Clements, F. E., Plant Succession. Carnegie Inst. Wash. Pub. No. 242, 1916.
5. Schuchert, C., The Making of Paleogeographic Maps, *Leopoldina*, Amerikaband, pp. 116–125, 1927.
6. Merriam, J. C., The Living Past. New York, Scribner, 1930.
7. Chaney, R. W., and Sanborn, E. I., The Goshen Flora of West Central Oregon. Carnegie Inst. Wash. Pub. No. 439, 1933.
8. Sanborn, E. I., The Comstock Flora of West Central Oregon. Carnegie Inst. Wash. Pub. No. 465, pp. 1–28, 1935.
9. McCornack, E. C., Thomas Condon, Pioneer Geologist of Oregon. Eugene, Univ. Oregon Press, 1928.
10. Merriam, J. C., A Contribution to the Geology of the John Day Basin, *Univ. Calif. Pub., Bull. Dept. Geol.*, vol. 2, no. 9, 1901.
11. Knowlton, F. H., Fossil Flora of the John Day Basin. U. S. Geol. Surv. Bull. 204, 1902.
12. Chaney, R. W., A Comparative Study of the Bridge Creek Flora and the Modern Redwood Forest. Carnegie Inst. Wash. Pub. No. 349, pp. 1–22, 1925.
13. Chaney, R. W., Geology and Paleontology of the Crooked River Basin, with Special Reference to the Bridge Creek Flora. Carnegie Inst. Wash. Pub. No. 346, pp. 45–138, 1927.
14. Chaney, R. W., Quantitative Studies of the Bridge Creek Flora, *Amer. Jour. Sci.*, 4th ser., vol. 8, no. 44, 1924.
15. Chaney, R. W., The Mascall Flora—Its Distribution and Climatic Relation. Carnegie Inst. Wash. Pub. No. 349, pp. 23–48, 1925.
16. MacGinitie, H. D., The Trout Creek Flora of Southeastern Oregon. Carnegie Inst. Wash. Pub. No. 416, pp. 21–68, 1933.
17. LaMotte, R. S., The Upper Cedarville Flora of Northwestern Nevada and Adjacent California. Carnegie Inst. Wash. Pub. No. 455, pp. 57–142, 1936.
18. Chaney, R. W., The Deschutes Flora of Eastern Oregon. Carnegie Inst. Wash. Pub. No. 476, pp. 185–216, 1938.

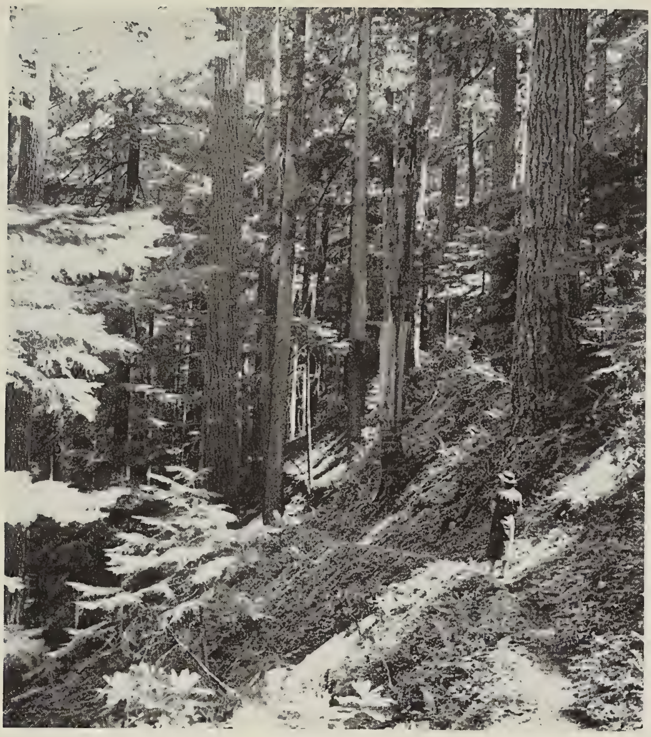


FIG. 1. Hemlock forest near the fossil locality above Cascadia, on the west slope of the Cascade Range.

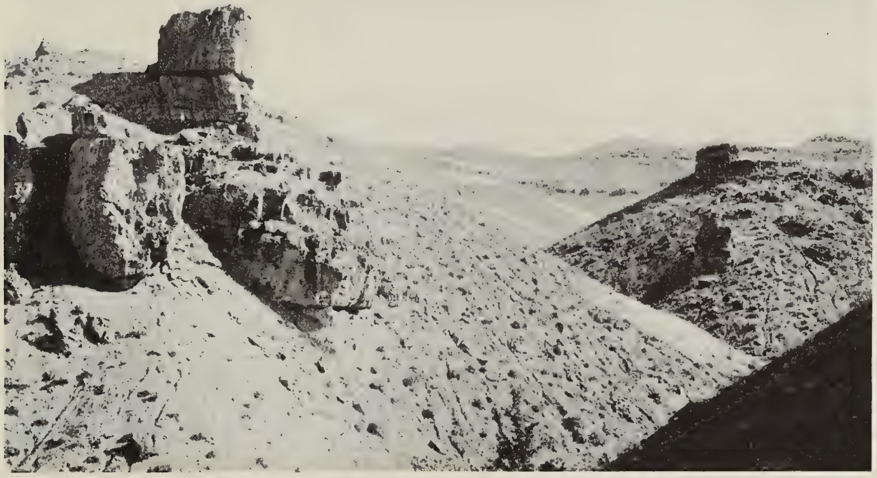


FIG. 2. Unforested hills on the east side of the Cascade Range near Clarno. A subtropical flora lived here during the Eocene, but today only a single juniper may be seen in the valley bottom.

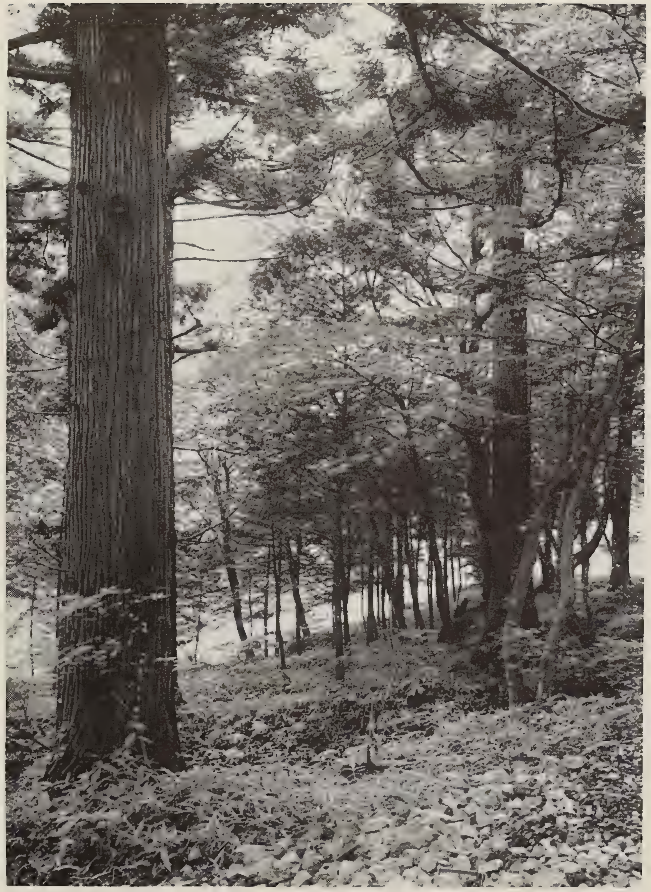


FIG. 1. Japanese cedars with an understory of oaks, maples, and hornbeams. This forest lies above Nikko at an altitude of 4000 feet.

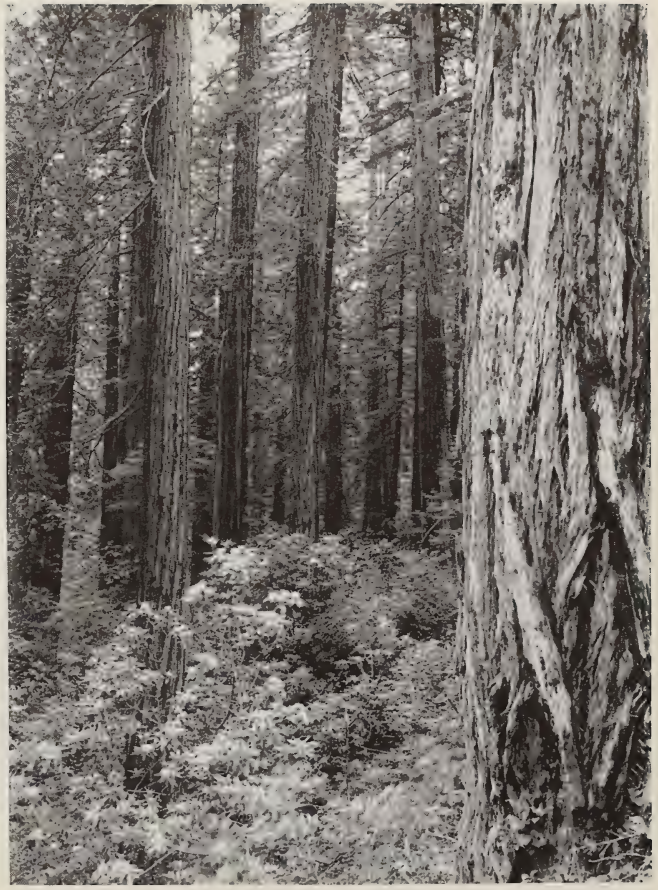
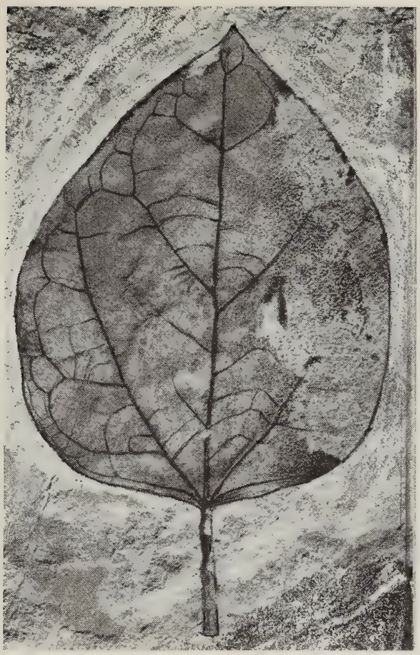
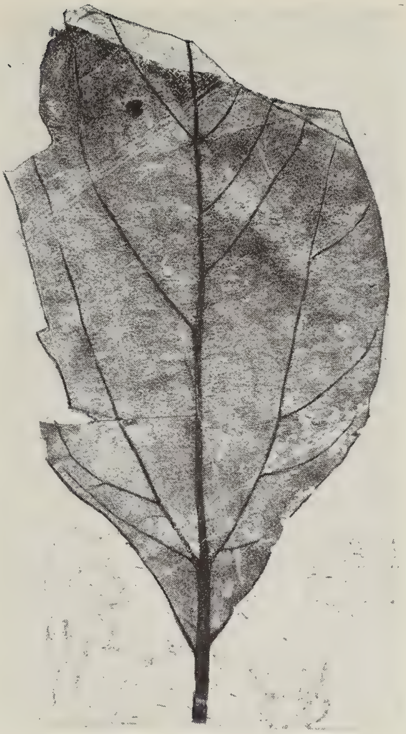


FIG. 2. Redwood forest near the California coast, in Del Norte County. Rhododendrons are the most conspicuous undershrubs. A forest like this ranged widely across Oregon during the Miocene.



Fossil leaf of catalpa from Miocene rocks near Rujada.



Fossil leaf of lancewood from Miocene rocks near Cascadia.

Mixed forest of warm-temperature and temperate species on a river border in southern Louisiana. With its avocadoes and palmettos, maples and elms, it resembles in composition the coastal forest of Oregon during the Miocene. Bald cypress takes the place of redwood.

ADAPTATION OF STRUCTURE TO FUNCTION IN WHALES

REMINGTON KELLOGG

Division of Mammals, United States National Museum
Research Associate, The Carnegie Institution of Washington

GEOLOGIC ages ago, whales and porpoises had ancestors that lived on dry land. For reasons not definitely known, some of these ancient relatives took to life in the water, possibly fresh water at first and salt water later. How and why did these land-dwelling mammals change, in the course of geologic time, into aquatic cetaceans, animals that must remain in the water or perish? This question has been a persistent puzzle since the days of the English zoologist John Ray (1671) and the Danish physician Thomas Bartholin (1654). Certainly there must have existed then as now along the banks of streams and in brackish marshes an abundance of water insects, shrimps, and other forms of aquatic life on which mammals could feed. Some of the mammals that lived in those ancient days along the banks of streams and small arms of the sea, noticing this abundance, may have learned to seek their food in the water, as other animals now seek similar food in like surroundings. Probably some of these shore-dwelling mammals were better adapted to swimming than others. Some of these gradually took to the sea, where they found a still more abundant supply of food. In some such manner as this the ancient ancestors of whales—small, active, hairy and warm-blooded—may have taken the first step in their transformation into creatures of the sea.

When one attempts to trace the geologic history of these ancient relatives of living cetaceans from those that lived on dry land to their descendants that hunted for food in the water near the shore, one finds many clues of absorbing interest in the fossil record. Today the fossilized bones of these ancient relatives exist in sedimentary strata such as sandy clays, diatomaceous earth, and various kinds of rocks. Many of these strata are now situated at elevations higher than the existing sea level and often at some distance from the present shores of oceans and seas. These sedimentary strata, when examined critically, are found to be ancient beaches, estuaries, and river deltas. The conditions under which these ancient cetacean relatives lived and died can be inferred to some extent from the composition and nature of the sediments in which their fossilized remains are interred, and from the animal and plant life preserved with them.

To decipher the evolutionary history of any group of organisms as recorded during geologic time, some knowledge of their nearest living relatives is required. The factors relating to the origin, structure, functions, and life history of extinct and living mammals, particularly the cetaceans, are so complex that no one investigator is competent to undertake a systematic examination of all the different problems that arise as the work progresses. One must therefore rely on the advice and investigations of colleagues who are specialists in particular fields of research. How essential these contacts with staff members of the Carnegie Institution of Washington, as well as with other research workers in other institutions, are in the appraisal of different aspects of problems as they arise may be better understood when one examines the adaptation of structure to function in a mammal that has become adjusted to habitual life in sea water.

These studies were commenced in the fall of 1919 at the University of California, where Dr. John C. Merriam, then chairman of the Department of Palæontology, had assembled cetacean material from Tertiary formations in California. At the outset it was planned that the writer should begin with the description of this accumulated material, but after some preliminary work it became apparent that any such study would require more adequate comparative Recent skeletal material than was then on hand. On January 1, 1920, the writer received an appointment in the government service at Washington, D. C. During September of the same year, Dr. Merriam suggested that this investigation should be continued under the auspices of the Carnegie Institution of Washington, through whose support it was possible to extend the scope of the investigation so as to comprise the broader field of the evolutionary history of the Cetacea. These studies were then continued, as opportunity permitted from regular duties at first as an employee of the Bureau of Biological Survey, United States Department of Agriculture, and later of the United States National Museum. In 1921, Charles W. Gilmore placed the fossil cetacean material of the Division of Vertebrate Palæontology, United States National Museum, at the writer's disposal, and in the ensuing years the privilege of studying cetacean material has been granted by the American Museum of Natural History; the Academy of Natural Sciences at Philadelphia; the Museum of Comparative Zoology, Harvard College; the Field Museum of Natural History; the California Academy of Sciences; the Los Angeles Museum of History, Arts and Science; the California Institute of Technology; the Department of Geology, University of Oregon; the Department of Geology, Millsaps College; as well as by a number of other American and European museums and institutions. The writer has at all times had free access to the extensive Recent cetacean skeletal material in the Division of Mammals, United States National Museum.

To digress a little from the main subject, it is interesting to record that John C. Merriam, while pursuing graduate studies under Professor Karl von Zittel at Munich, Germany, in 1892, met Professor Eugene W. Hilgard, who told him about the abundance of bones of ancient whales in certain parts of Mississippi and stimulated his interest in these extinct Eocene mammals. From 1858 to 1873, Professor Hilgard had been employed in Mississippi as state geologist and as a professor in the State University, and from 1875 to 1904 as professor of agriculture at the University of California. After obtaining his doctor's degree at Munich, Dr. Merriam returned to the United States and, on the way to his home in Iowa, visited the World's Fair at Chicago. There he aided Professor Thomas C. Chamberlin in gathering for the University of Chicago some of the materials exhibited at the Fair. Subsequently, Dr. Merriam remained for two quarters at the University of Chicago. Dr. Merriam promised to remain for a time if some way could be found to secure funds for collecting some of this ancient whale material in the Gulf Coastal Plain. Some weeks later he was invited to have dinner with President William Rainey Harper of the University of Chicago and on that occasion repeated the same general remarks he had made to Professor Chamberlin. Funds not appearing, Dr. Merriam later accepted an instructorship in palæontology at the University of California. This plan for collecting in the Gulf Coastal Plain was then laid aside. No further steps were taken until 1928, when, after a conference with Dr. Merriam, it was decided that the writer should undertake an investigation of the Archæoceti.

In this summary an attempt will be made to review the results obtained from a study of fossil and living cetaceans and to determine what the evidence so obtained signifies in the light of other related studies made during the past decade.

Value of Fossil Cetaceans for Determining Age of Sedimentary Strata

In the earlier stages these studies were largely descriptive, but as they progressed the importance of fossil cetaceans for geological correlation became more apparent. This led to an examination of the occurrences of fossil cetaceans in all parts of the world. In this inquiry, it was necessary to rely on the advice of competent geologists who had been made familiar with the stratigraphic positions of the respective horizons by actual field experience. This required extended conferences and correspondence with many geologists and invertebrate palæontologists. A preliminary discussion (Kellogg, 1924) of the available evidence has been presented as part of the "Symposium on the Correlation of Tertiary Formations, etc." Although the correlation possibilities of fossil cetaceans were never lost sight of, it soon became apparent that this phase of the investigation must necessarily await the accumulation of larger collections from Tertiary formations in North America. As a collateral investigation, the recorded occurrences of migrating whales in the several oceans were assembled (Kellogg, 1929). These observations show that the Recent whalebone whales make seasonal migrations from tropical calving grounds to the feeding grounds in the colder waters of Arctic and Antarctic Seas, and that many of the larger cetaceans are practically cosmopolitan. There is no reason to assume that the cetotheres, the Tertiary precursors of the Recent whalebone whales, did not follow similar migration routes. Hence their remains should occur in geological horizons of corresponding age on the old shores that bordered these oceans. More intensive collecting of cetacean material during the past decade in the North American Miocene marine formations tends to corroborate this assumption.

MODIFICATIONS IN BODILY STRUCTURE RELATED TO AQUATIC REQUIREMENTS

The interpretation of exceptional structural conditions, which were discovered in the course of these taxonomic studies on the cranial and skeletal remains of extinct cetaceans, required careful examination of the skeletons of practically all of the Recent cetaceans. The writer had the good fortune to be able to discuss many of these details with Gerrit S. Miller, Jr., who has published an account (1923) of the telescoping of the cetacean skull. These stimulating discussions were of inestimable value in attempting to arrive at reasonable explanations for puzzling structural details. Likewise the investigations of Professor A. Brazier Howell (1930) on the functional anatomy of mammals specialized for swimming or diving, which were carried on largely in the Division of Mammals, United States National Museum, emphasized the astonishing modifications in bodily structure that were related to aquatic requirements. While the external form of the cetacean was "streamlined" for aquatic life and some structures were apparently simplified, no fundamentally new structures were added to the internal anatomy, but only specializations of many of the old ones. Of all mammals extinct or living of which we have any knowledge, the Cetacea are obviously the most highly modified. The following brief résumé is given to indicate the complex nature of the anatomical alterations and physiological adjustments that were effected in the evolutionary process of changing a terrestrial mammal into one adapted for a pelagic existence.

All the known cetaceans are air-breathing warm-blooded mammals, generally having pointed heads, torpedo-shaped bodies, finlike forelimbs, and horizontal caudal flukes. Included in this group are the extinct zeuglodonts (Archæoceti), the whalebone whales (Mysticeti), and the toothed whales (Odontoceti).

Although we can only surmise the incipient modifications that facilitated the first steps in the transformation of a terrestrial progenitor into a whale-like animal adapted for aquatic life, it is obvious that the functioning of the entire cetacean body is now conditioned by adjustments to an exclusively marine habitat. Among the alterations necessitated by an exclusive marine life are the perfection of a mechanism for propulsion in water, a physiological adjustment to drinking salt water, the remodeling of the dentition for a specialized kind of food gathering, the adjustment of the viscera, particularly the digestive tract, for the assimilation of a specific kind of food mixed with quantities of sea water, modifications in the nostrils and throat to prevent the entrance of water into the lungs, thermal adjustments to eliminate rapid loss of body heat, and adjustments in the reproductive organs to make pelagic breeding possible. Some of the more important details in this extremely complicated process of transformation may be summarized as follows:

To cope successfully with the conditions imposed by habitual life in sea water, the physical equipment for propulsion must be profoundly altered. This swimming equipment must be sufficiently efficient to permit sustained progression for considerable periods of time. In the case of whales, we observe that the motive power was transferred from the limbs, which originally served as ambulatory organs, to the hinder end of the body, resulting in the development of horizontal caudal flukes, which became the propelling mechanism. Any animal which employs its tail for propulsion must have an anterior steering mechanism. Otters, pinnipeds, and *Potamogale* employ the neck for steering purposes, but the cetacean neck is so short that it cannot be so used. Thus a separate mechanism for steering and balancing is required and the most available structure is the forelimb. The pectoral flippers of cetaceans bear little resemblance to the forelimbs of land mammals, for they are overlain with blubber and are fashioned like a paddle with margins hardened by fibrocartilage. With the passage of time there has come the complete envelopment of the fingers and the loss of nails, the wrist bones have lost their original shape, and the entire forelimb has been transformed into an elastic plate movable only at the shoulder joint. A detailed knowledge of this sort of adaptive modification can only be obtained by observations on swimming cetaceans, by an examination of the bones comprising the forelimbs of extinct and living cetaceans, and by dissections of fresh or embalmed bodies. In this I have been privileged to observe the progress of dissections made by Professor A. Brazier Howell on the Chinese porpoise *Neomeris* (1927) and other cetaceans.

Although a vestigial pelvis is present in most toothed whales, no trace of the hind limb bones has been recognized. We are fairly certain, however, that hind limbs were present in the progenitors of all Cetacea because the archæocetes and the whalebone whales have one or two vestigial limb bones buried deep in the flesh of the pelvic region.

Physiological Adjustments to Drinking Salt Water

The physiological adjustments that were brought about to obviate thirst for fresh water during the transition stages to an offshore marine life offer a fertile field for speculation. Myers (1920, p. 142) and Sudzuki (1924, p. 426) have shown that the salt content of whale blood is strikingly high as compared with that of land mammals. The bodies of all animals that serve as food for cetaceans contain considerable salt and consequently some of this salt is absorbed as non-digestion products into the blood stream. Portier (1910, pp. 206–207) concluded that although the greater part of this salt was immediately eliminated through the action of the kidneys, the periodic injection of salts into the blood stream finally resulted in making the body liquids isotonic with sea water. It is not known by what means Recent whales are able to avoid assimilating excessive quantities of salts from their food or from the sea water that they swallow. It is evident that some sort of mechanism exists to control the taking up of salts, so that Recent Cetacea are able to take from the sea necessary quantities of water isotonic with the blood and pass it immediately through the kidneys. The kidney of the adult whalebone whale is made up of a number of little kidneys called "renculi," each being complete with cortex, medulla, pelvis, and calyx, and its own arterial and venous blood system. Ommanney (1932, p. 457) calculated that the total number of renculi in the whole mass of each kidney of a finback whale was between 5998 and 6372 and that the length of the kidney was dependent on the length of the cetacean. This aggregation of little kidneys closely packed and contained within a common envelope may provide a more effective mechanism for the elimination of the excess salts.

Food, Food Getting, and Digestive Alterations

Food getting would necessarily play a prominent part during the initial stages of adaptation to an aquatic life. During that period a flesh eater would feed on fish or on prey which did not possess some sort of effective protective covering, such as a hard shell or spiny armature. By the time the animal had become so completely adapted to water that it had lost all means of progression on land, it would be forced to subsist on prey small enough to be either swallowed whole or held in the mouth while it is being chewed. Adaptive transformations came about during which at first some and later all of the teeth lost their function of crunching and tearing meat and became remodeled for the grasping and holding of prey, a function that had been lost by the remodeled forelimbs. When the animal attained a large size, the food requirements increased in proportion. Feeding periods would be lengthened. If the animal persisted in feeding on small prey, ultimately it would reach the limit of its endurance unless corresponding changes in the bodily metabolism were brought about. Since every animal is a physiological unit it therefore follows that any impairment of an essential structure will affect to some degree, temporarily at least, the metabolism of the whole body. Mechanically necessitated structural changes are thus closely tied up with the physiological requirements of the body.

Thus one observes that the trigonal style of cheek teeth possessed by the earliest known Eocene archæocetes became converted in later members of the group into high-crowned cheek teeth with serrated cutting edges. The latter style of cheek tooth is likewise possessed by the earliest known types of odontocetes and persisted in some genera until the close of the Miocene period. The precise details of this general adaptive trend in which the dentition was remodeled for a specialized kind of food gathering are not as yet fully understood. Nevertheless, most of the living odontocetes and many of the extinct forms possess a dentition characterized by the multiplication and reduplication of cylindrically shaped teeth.

Further adjustments to unexpected conditions associated with the procurement of food subsequently appeared in the course of the geological history of cetaceans. Embryology furnishes the evidence that the progenitors of the mysticetes were provided with teeth. As many as forty or more minute teeth with pin-like crowns may be found hidden in the gum on a maxillary of a fetal mysticete, but they are soon resorbed. When the finback fetus grows to a length of about 10 feet, the development of papillæ along the outer border of the upper jaw increases and spreads. These papillæ subsequently develop into a series of crosswise-placed corneous plates, the baleen. The late Oligocene and early Miocene cetotheres, precursors of living whalebone whales, had already lost their teeth and acquired the two rows of transversely placed blade-like plates of baleen that were attached to the roof of the mouth and served as a seine when the animal was feeding. Small crustaceans and fish constitute the food of living whalebone whales; the cetotheres no doubt subsisted on similar types of food. An abundance of soft-bodied crustaceans and small fish provided ample and readily obtainable food, which required little or no mastication, but which did require some structure adapted for catching such minute prey. These plates of baleen with internal fringes of intermatted bristle-like fibers form a strainer remarkably well adapted for impounding swimming crustaceans and small fish which are scooped up when the whale plunges forward with open mouth. The remodeling of the mysticete skull seems to have been profoundly influenced by the development of baleen.

In the living mysticetes, the krill is engulfed in the oral cavity along with a large quantity of sea water, most of which necessarily must be expelled by the tongue assisted by the specialized gular musculature before such small prey is swallowed. The odontocetes likewise procure their food while swimming; it varies, according to species, from fish to squid, octopus, and sting ray. Hence, in all cetaceans, the viscera, particularly the digestive tract, must be altered for the assimilation of a specific kind of food mixed with quantities of sea water. As has already been noted, the physiological adjustments necessary for the utilization of sea water by a warm-blooded mammal have not been adequately investigated. Cetaceans have a stomach that is subdivided into several chambers. Fourteen of these compartments are recognizable in the ziphioid stomach, while the stomach of balænopterine whales has only four chambers. In the majority of the toothed whales, the first division of the stomach appears to be a dilation of the œsophagus, but in the ziphioids the structure of the stomach indicates that the first division may be missing and that the existing divisions are all true gastric chambers. These successive chambers are sharply separated from one another by septa in whose center are the communicating openings. The divisions of the whalebone whale stomach are less sharply differentiated from one another and the communicating openings are wide. A whalebone whale consumes a large quantity of shrimps or other marine animals at a single meal. The mass of krill (*Euphausia superba*) in the stomach of one blue whale weighed more than a ton (Peters, 1938, p. 124). This food must be acted on by the gastric juices and altered so that it becomes soluble. The multiple divisions of the cetacean stomach probably serve to divide the food mass into parts that are small enough to enable the gastric juices to accomplish their share of the digestive process. The complicated stomach is connected with a relatively long intestinal tract. Although the entire intestinal tract of full-grown porpoises measures from 13 to 16 yards in length, a 17-foot, immature little piked whale (*Balænoptera acuto-rostrata*) dissected by John Hunter (1787, p. 407) had a small intestine 28.5 yards in length and a large intestine 2.75 yards long. The increased surface needed for adequate digestion and absorption in the baleen whales has been provided for in part by this lengthening of the intestine and in part by a great increase in the number of minor loops in "Meckel's tract," which is suspended at the circumference of the mesentery. Thus the digestive tract seems to be designed so that the food undergoing digestion is brought in contact with an increased area of absorptive surface so that the maximum amount of nutritive materials can be extracted.

Modifications in Nose, Throat, and Lungs Related to Respiration

Adjustments and modifications in the structures which function in the process of breathing were necessary to permit feeding. Water would surely get into the lungs during swimming with a wide-open mouth were it not for the fact that a whale's narial passages do not open into the throat like those of land mammals. A continuous tube between the narial passages and the trachea or windpipe is formed by an elongation of the arytenoid cartilage and the epiglottis which fits into the soft palate. In this way a whale, when feeding, can swim with its mouth wide open without any likelihood of being strangled.

In order to obtain complete success as an air-breathing water-living mammal, alterations are necessary in the mechanism for closing the nostrils. Closure of the nostrils is accomplished in an entirely different manner in the two groups of living whales. The toothed whales have one external nostril, while the whalebone whales have two external nostrils.

In the case of porpoises, the narial cavity is divided into a series of pockets or sacs into which the narial passages open, and between these sacs there is a series of valves controlled by muscles which apparently relax when the outside water pressure on the flap of fibroelastic cartilage covering the external nostril is removed on emergence from a dive. With the contraction of a special set of muscles the external orifice of the blowhole is closed, compressing the spiracular cavity and premaxillary diverticula, and as a result the air compressed in these cavities exerts sufficient pressure to force the plugs firmly into the dorsal entrances to the bone-enclosed narial passages. To the late Dr. Ernst Huber the writer is indebted for a much closer understanding of the odontocete nostril-closing mechanism, gained in part while watching his preliminary dissections at Hatteras, North Carolina, and subsequently from his published observations (1934) on the facial musculature of the bottle-nosed porpoise, the narwhal, and the blackfish.

In the case of the whalebone whales we observe that the tubular narial passages slope forward and upward, and are furnished with spiral folds which fit into one another when pressure is exerted from the outside. When feeding, the whalebone whale makes periodic submergences, since the food sought is found according to conditions at varying levels between the surface and a depth of 100 m. When a whale comes up after a dive its head is the first part to appear above the surface. Expiration does not take place before the water surface has been broken. The blast of expiration or "spout" lasts from 0.6 (Laurie, 1933, p. 370) to 1 second (Walmsley, 1938, p. 170). Immediately after the expiratory blast the lungs are refilled, the blowholes being opened widely and protruded upward, and the air drawn in. The elevation of the blowholes probably prevents a wave from slopping over and filling the nasal passages with water. Inspiration lasts from 0.9 to 1 second, the nostrils close, and then the whale prepares to dive. The whale remains at the surface from 1.5 to 2 seconds before submerging for about 15 seconds, after which the respiratory act is repeated. According to Walmsley (1938, p. 170) the whale makes about ten of these shallow dives before making a prolonged submergence of 7 to 10 minutes. Although the closing mechanism is constructed along different lines in the two groups of cetaceans, the nostrils are effectually closed whenever an animal of either kind is submerged.

Thermal Adjustments That Prevent Rapid Loss of Body Heat

Thermal adjustments are necessary to prevent loss of body heat during constant immersion in water. Since the capacity of water for conducting heat is 27 times as great as that of air, whales would lose a vast amount of body heat were it not for the presence of a blanket-like heat-retaining layer of fat over the whole body, and the absence of sweat or sudoriferous glands. The lack of sweat glands, however, eliminates one method of body temperature regulation from consideration. There remains, however, the possibility that whales may modify body temperature either by variation in the blood supply to the skin or by control of the basal metabolic rate. Whales in tropical waters need this protection as much as those in temperate or colder waters, for the individual body temperatures of whales always average higher than the temperature of the water in which the animals are immersed. The body temperature of finbacks and blue whales is given by Laurie (1933, p. 366) as slightly more than 95°F., and that of the common harbor porpoise (*Phocæna*) is reported to be 96°F. (White, 1891, p. 385). Whalebone whales are generally present on the feeding grounds in the colder waters for more than half of the year. The temperature of the surface water in either the Arctic or Antarctic is rarely above 39°F. and frequently below 32°F. With the exception of the time required to migrate from and to the colder waters, the remainder of the year is spent by many whales on the tropical calving grounds, where surface water temperatures range from 80° to 84°F.

Adaptations That Make Aquatic Breeding Possible

It might be expected that adjustments would have been made in the cetacean reproductive tract to prevent expulsion of the fetus from the womb when the body of the mother is subjected to considerable pressures during prolonged submergences. Meigs (1849, p. 270) asserted that the facultas retentrix has been fortified by an apparatus present in no terrestrial mammal. This apparatus is described as consisting of a lower cervix and os externus in the canal of the vagina and an "upper cervix and os [internus] enclosed in the canal of the cervix" (Meigs, 1849, p. 267). More recently, Dr.

George B. Wislocki of Harvard University Medical School undertook an investigation (1933) of the gravid reproductive tract of a harbor porpoise and described an elaborate set of folds at the junction of the uterus and vagina. Nevertheless, Wislocki (1933, p. 96) concluded that the question of whether the cervix should be regarded as a highly developed and specialized structure must await further comparative morphological studies. Ommanney (1932*a*, p. 406 and fig. 12) has described the reproductive tract of a finback which has a cervix 0.9 cm in length with a comparatively narrow os internus and a wider os externus leading into the vagina. Furthermore, in the wall of the fore part of the vagina there are three circular folds that form backwardly projecting funnels, reducing this structure to extremely narrow dimensions.

The body of the newborn cetacean most certainly is disproportionately large as compared with the young of other mammals and also unusually advanced in development. At birth the young cetacean immediately becomes a free-swimming individual, dependent on its own exertions not only to escape from being drowned, but also to suckle its mother, and to avail itself of such protection as the mother may be able to render against predatory enemies. The mental and physical capacities of the newborn cetacean must be developed sufficiently to cope with the aquatic environment, since it cannot be hidden like the young of land-dwelling mammals until such time as it is physically able to make concerted use of its faculties.

The normal manner of suckling young has been modified, for the mammary glands of whales have large sinuses or reservoirs in which the milk collects. By forcible contraction of associated body muscles these reservoirs of milk are emptied by way of the teat into the mouth of the calf, so that nursing is a matter of a few moments.

Modifications in the Eye

To obtain complete success as an air-breathing aquatic mammal, adjustments were necessitated in the eye, ear, and nose. The functioning of the respiratory and circulatory apparatus was likewise affected by the change to the new environment. Probably the basal metabolism underwent corresponding alterations. These and other adaptive details can now be considered.

A number of investigators have pointed out that the eyes of many of our living land mammals do not possess the power of accommodation. It is therefore not unlikely that similar conditions prevailed among the mammals of the past. A normal or emmetropic eye is one in which parallel rays of light reaching the eye through the air are brought to a focus on the retina. Now if a land mammal with an eye of this type should be transferred to a pelagic habitat, its vision would be impaired, for only the outlines of objects viewed under water would be recognizable, because the light rays reaching the eye through sea water would be focused beyond the retina. Even with the maximum accommodation, in which the anterior and posterior surfaces of the lens have the same radius of curvature, parallel rays of light reaching the eye through sea water would be focused beyond the retina. Unless the axis or height of the lens was increased and the refractive indices of the liquid media surrounding the lens were altered slightly, the eye would function imperfectly when immersed in sea water.

We have unquestionable evidence that the eyes of cetaceans are adjusted to the specific functions they have to perform, even though they lack the power of accommodation. Inasmuch as the refractive index of sea water is not the same as that of air, the eyes of whales must have needed alteration in order to function when immersed. There is so little difference between the refractive indices of sea water and of the refracting fluid in the anterior chamber of the eye that the cornea is dioptrically inoperative when the eye is submerged, and the dioptric apparatus is thus reduced to the lens. To compensate for this, however, the curvature of the anterior and posterior surfaces as well as the axis or height of the lens, and the refractive indices of the liquid media have been corrected so that light rays passing through sea water are brought to a focus on the retina.

Under atmospheric conditions the eyes of most whales are myopic, that is, the light rays are brought to a focus in front of the retina. Thus when a whale tries to view objects above the water level it is troubled not only with astigmatism because of differences in curvature of the cornea in the two directions, but also in being unable to change the curvature of the lens for optical accommodation because of defective ciliary muscles. Dr. Fred E. Wright of the Carnegie Geophysical Laboratory kindly checked the above computations, which are based on the refractive indices of the lens and of the liquid media of the whalebone whale eye.

The chemical adjustments of the whale eye are equally interesting. Constant chemical irritation resulting from contact with sea water necessitates certain adjustments to protect the eye from inflammation or even more serious injury. The tear gland of the whale does not secrete a watery fluid, for the eye is no longer in danger of drying up. It exudes a greasy substance to protect the cornea from various chemical substances found in sea water. Furthermore, in a number of whales the conjunctiva or mucous outside coating of the eye has been replaced by cornified epithelium. Since the whale eye is subjected to almost constant immersion in sea water of 3 to 4 per cent salt content it does not appear unreasonable to assume that these details represent adjustments to chemical irritants. Since the cornea of the whale eye is in direct contact with water, one might expect that the eye would be protected in some way against continuous cooling below normal body temperature. This seems to have been effected in part by lowering the freezing temperature of the liquid media in the entire eye and in part by accelerating the inflow and drainage of the liquid in the eye's anterior chamber.

Modifications in the Organ of Hearing

Life in the water necessitated a remodeling of the organ of hearing so that it would function under the changed conditions. We know that whales at one time possessed some sort of external ear, for some porpoises have a vestige of the auricular cartilage buried beneath the skin, and the two outer ear muscles are still retained by the baleen whales. In many of the porpoises the outer orifice of the external auditory tube is so minute that it will not permit the passage of the lead of a pencil, although the lumen expands in a trumpet-like manner as it approaches the eardrum. The outer orifice of the external auditory tube of a humpback whale is often concealed by a fold of skin. In one of these whalebone whales the external auditory tube measured 21 inches in length, but was completely closed near the middle for a distance of 3 inches. The tympanic membrane of a mysticete is in many respects the most remarkable structure connected with the organ of hearing. It is a finger-shaped structure, about 3.5 inches in length, and is soft and flexible in the fresh state. The complete closure of the external auditory tube must have been accompanied by corresponding physiological adjustments to prevent the accumulation of an excess amount of wax in this blind tube.

The initial stages in the transformation of an organ of hearing designed for the perception of air-borne sounds to one adapted to receive water-borne sounds are unknown, for the earliest archæocetes yet found have the osseous portions of the ear as highly developed as any of the living whales. The remote ancestors of the living whales must have had the usual type of normal eardrums; but in the course of geological time the osseous portions of the inner ear have undergone certain changes which brought about a diminishing utility for the tympanic membrane.

Whales have acquired an organ of hearing in which resonance must play an important part, for the malleus is rigid, with its anterior process fused with the outer lip of the bulla; the tensor tympani muscle, whose function in land mammals is to make taut the eardrum, is vestigial; and in some species the stapes is almost immovable in the fenestra ovalis. The tympanic bulla is the relatively dense and heavy sounding box fastened to the periotic by two thin pedicles which can be set in vibration. Vibrations set up in these pedicles produce a corresponding amount of motion in the malleus, whose anterior process is likewise fused with the bulla between these pedicles, and it in turn transmits these vibrations to the incus and stapes. Another mechanical arrangement is thus acquired to take over the function ordinarily performed by the eardrum. Water-borne sound vibrations transmitted to the air contained in the tympanic bulla cause it to function as a sounding box, and its vibrations reach the cochlea by way of the ossicular chain and the vestibule.

Through the interest of Dr. Lewis H. Weed, a grant was made in February 1928 from the Linton Fund of the Johns Hopkins Medical School for members of the staff of the Department of Anatomy, including Drs. George B. Wislocki, Ernst Huber, Orthello R. Langworthy, A. Brazier Howell, Adolph H. Schultz, and the present writer to visit the porpoise fishery conducted by Howard I. Wordell at Hatteras, North Carolina, for the purpose of assembling suitable material for more detailed studies relating to the general problem of adaptation. The brains of a number of bottle-nosed porpoises (*Tur-*

siops truncatus) were secured, one of which was dissected by Dr. George L. Streeter in the Carnegie Embryological Laboratory. He found that these porpoises have a highly developed central nervous system for the reception and disposal of cochlear stimuli, which stands out in contrast to their less well-developed optic mechanism. The auditory colliculus is four times the size of the corresponding optic colliculus. More detailed studies by Dr. Langworthy (1932; Ries and Langworthy, 1937, p. 34) confirmed the assumption that aquatic adaptation has produced marked changes in the central nervous system and that the chief factor contributing to the differentiation and growth of the cerebral cortex is the great influx of sensory impulses, particularly auditory stimuli from the cochlear nerve.

Modifications in the Olfactory Structures

Chemical senses, like those of smell and taste, which are closely associated physiologically, require organs which differ considerably in structure and relations. The sense of smell plays an important part in the welfare of most mammals, but whales apparently found less and less need for olfactory structures.

The endocranial casts of some of the upper Eocene archæocetes are unusually complete. They show not only the general conformation of the brain and its nerves, but also the cavities and sinuses in the region of the narial passages. It is obvious that the sense of smell was still of considerable importance to these archæocetes, as is shown by the development and peculiar elongation of the olfactory peduncles, which are attached to the ventral surfaces of the cerebral hemispheres in the characteristic mammalian fashion. The rather smooth cerebral hemisphere, which is primarily an olfactory receptive mechanism, is quite different in appearance from the convoluted cerebrum of living odontocetes and mysticetes. An examination of the endocranial casts of archæocetes and extinct porpoises from successive geological stages indicates that the loss of the sense of smell was gradual and that this loss was accompanied by an excessive development of the trigeminal apparatus, which in a pelagic mammal provides much more information concerning food and enemies.

The whalebone whales have olfactory structures, but the retention of the sense of smell may be due in a larger measure to the actual mechanical construction of the skull than to the need for such organs, for the olfactory structures seem to be adapted for smelling through the medium of air as in terrestrial mammals. Water never comes in contact with the sinuses in which the olfactory nerves are distributed and no odors other than those conveyed by water could be recognized by whales that are feeding below the surface. We are therefore at a loss to conceive how whales may smell through the medium of water.

The living toothed whales and porpoises have lost their olfactory organs, though such structures have been found in embryos. Mechanical changes in the relations of the component parts of the odontocete type of skull appear to have restricted at first and finally prevented the physiological functioning of the olfactory apparatus. These changes have hastened the final reduction of the dwindling olfactory apparatus of the toothed whales. On the basis of skulls of Miocene porpoises, one would be led to conclude that the passages for the olfactory nerves were at first reduced and later entirely closed by the upward and lateral expansion of the ectethmoids. The general course of the reduction of the olfactory organs can be demonstrated also by comparison of brain casts of fossil porpoises from successive geological stages.

Adjustments to Pressure during Prolonged Submergence

The hydrostatic adjustments that must have been effected in the whale's body are even more complex and puzzling. Notwithstanding the fact that the maximum depth reached by any whale is still in dispute, there exists some evidence that is incontrovertible. In April 1932, P. E. Harne, captain of the cable repair ship *All America*, while investigating the cause of an interruption in the service of the submarine cable between Balboa, Canal Zone, and Esmeraldas, Ecuador, hoisted to the surface a dead 45-foot male sperm whale from a depth of 3240 feet (540 fathoms) off the coast of Colombia. The submarine cable was twisted around the lower jaw of this animal and an additional 30 fathoms of cable were wrapped around one flipper, the body, and the caudal flukes (*Inquirer and Mirror*, Nantucket Island, Massachusetts, April 9, 1932). External pressure in water increases by one atmosphere (14 pounds per square inch) for each 10 m of depth. At a depth of 988 m (540 fathoms) this sperm whale must have been subjected to a pressure of approximately 1400 pounds per square inch of body surface. A similar incident, but at a lesser depth, occurred in Alaskan waters. A cable repair ship found that a 50-foot humpback had snarled a loop of submarine cable around its rostrum at a depth of 119 m (65 fathoms) 10 miles off Sitka (*Scientific American*, vol. 92, p. 447, 1905). At that depth the humpback would have to withstand a pressure of 13 atmospheres or 182 pounds per square inch.

The hide of a whale is endowed with certain structural features which afford the maximum elasticity, and conversely may retard to a certain extent increasing compression on the thorax. The fibrous blubber has tremendous strength and tenacity, for it will resist a strain of several thousand pounds. Finbacks have been known to tow a ship when the prongs of the harpoon were merely caught in their blubber. It must be borne in mind, however, that the blubber does not act as a rigid or semi-rigid sheath. Though fairly hard, it is flexible and quite pliable, and in the case of balænopterine whales is slack ventrally in the region of ventral grooves from chin to umbilicus.

The whale eye has been adapted to withstand varying pressures as well as the impact of the water when swimming at the surface. It is a well-known fact that a vault designed to withstand heavy pressures has to be built so that the directions of the strains, stresses, and pressure forces fall within the abutment, and in the case of the whale's eye this abutment is the enormously thickened sclera. Fluctuations in hydostatic pressure result in corresponding changes in intraocular pressure.

In the case of any deep-diving mammal, adaptations for hydrostatic pressure must be effected not only in the eyes, ears, and nose as described in preceding paragraphs, but also in the respiratory and circulatory systems.

Adaptation of structure to function is no less complicated in the respiratory tract than in the other organs heretofore enumerated. The unusual quantity of elastic tissue and musculature in the lungs, the thickness of the pulmonary pleura (Haynes and Laurie, 1937), the presence of circular and closed cartilages and the absence of membranous portions in the trachea and bronchi, the presence of a cartilaginous armature extending to the terminations of the bronchial tree, the existence of numerous valves or muscular sphincters in the lumina of the respiratory bronchioles and a complete lining of flattened respiratory epithelium (Wislocki, 1929, pp. 57, 58), the direct communication of the respiratory bronchioles through sphincter openings into the alveolar air sacs, and finally the large size of the alveolar air sacs and the thickness of the septal walls, all point to a respiratory mechanism highly specialized to meet the increased pressure on the surface of the submerged body.

In addition to alterations that have been described (see page 658) for the nasopharynx and the nostrils to facilitate expiration and inspiration and to close the external openings of the air passages during submergence, there are other modifications that appear to be related to respiratory requirements. The neck has been shortened and rendered relatively immobile, the trachea is short, and the barrel-shaped configuration of the thorax and the oblique position of the diaphragm give added capacity to the thorax. Flexibility in the dorsal series of vertebræ, associated with a relatively short sternum, has been interpreted as an adaptive alteration for resisting the temporary effects of compression on the thorax during submergence.

The ability of the heart and lungs to function under the influence of high pressures of water acting on the thoracic wall involves, as has been pointed out, alterations not only in the trachea and lungs but also in the thorax and diaphragm. The ability to suspend breathing while submerged depends in part on adequate oxygenation of the blood, the release of nitrogen when pressure is removed, and the disposal of carbon dioxide. The purely chemical side of respiration can only be surmised from the more or less incomplete anatomical data and the gas analyses of blood and other body liquids.

Instead of having a continuous ventilation of the lungs like land mammals, the whale has to depend on an intermittent filling and emptying of the lungs as described on page 659. Widely divergent opinions have been advanced by competent physiologists and anatomists to explain the respiratory mechanism during submergence. It will suffice to mention only two of these views. One assumption is that the submerged whale is dependent for its oxygen supply solely on the air inhaled into the lungs. According to those holding this view, the whale takes a series of short breaths, thoroughly ventilating the respiratory air sacs before filling its capacious and very elastic lungs with air for a prolonged submergence; the whale returns to the surface when the oxygen in this air is nearly exhausted.

The other assumption is that the whale stores up an oxygen supply in its blood and fatty tissues by the usual expiration and inspiration between a series of ten or more shallow submergences preparatory to a more prolonged submergence. The submerged whale, after exhausting the air in its lungs, then draws on the oxygen stored in the blood and in the fatty tissues surrounding the rather extensive retia mirabilia.

The depth to which whales dive is a matter of considerable interest physiologically, since the external pressure encountered during submergence should affect all the vital processes if it is communicated to the lungs and the air contained in them. According to the available evidence (see page 666) it is a fair assumption that the larger whales do at times submerge to a depth of 100 m in search of food. At a depth of 100 m the whale's body, including the thorax, will be subjected to a pressure equivalent to eleven atmospheric pressures (154 pounds per square inch), for the water pressure increases by one atmosphere (14 pounds per square inch) with each additional 10 m of submergence.

The assumption that the lungs of whales can be compressed by hydrostatic pressure to the extent of more than one-quarter of their capacity has been advocated by Laurie (1933), Damant (1934), and Ommanney (1935), and questioned by Krogh (1934*a*). Wislocki (1929, pp. 65, 66) holds that the closure of the sphincters when inspiration is completed, in conjunction with the cartilaginous armature, is a mechanism adapted for impounding the air in the air sacs and respiratory bronchioles, for resisting the gradual collapse of these air spaces, and for maintaining the air in contact with the respiratory mucosa during the interval when the water pressure on the thorax is communicated to the lungs. Furthermore, the passageways between the lungs and the exterior are not at all times open as they are in land mammals, since the narial openings are closed tightly during submergence, and the sphincters in the respiratory bronchioles are adapted to holding the air. The thorax is nearly circular in cross section and the fibrous blubber overlying the intercostal and abdominal musculature may resist compression to some extent. On the other hand Irving (1935, p. 560) has pointed out that if intrathoracic pressure remains low, while external hydrostatic pressure increases, the viscera, blood, and lymph will be forced into the thorax.

Assuming that the volume of gases in the lungs will be compressed in direct proportion to the increase of external pressure on the thorax of a whale, the partial pressure of the oxygen in the air contained in the lungs will be raised. Although the partial pressure of oxygen in air at sea level is about 150 mm of mercury, a partial pressure of 60 mm of mercury is sufficient to keep the blood of most mammals above 80 per cent saturation with oxygen. According to the computations made by Laurie (1933, p. 371) the partial pressure of oxygen in a whale's lungs will be equivalent to 1500 mm of mercury at a depth of 100 m and thus at that depth the air may become ten times as depleted of oxygen before inadequate oxygenation of the blood sets in. The oxygen in the lungs would be at a partial pressure of only 6 mm of mercury when the whale returned to the surface. Under these conditions the whale is able to make the fullest possible use of the inspired air.

Although allowance has to be made for the gas analyses given by Laurie (1933, pp. 396–397), owing as he says to faulty apparatus, the average oxygen capacity of 40 samples was reported as 213 cc of O_2 per liter of blood, and of 17 samples as 141 cc of O_2 per liter of blood. Taking the oxygen capacity at 213 cc, the 8000 ± l (8000 ± kg) of blood of a 122,004-kg blue whale would absorb 1704 l of oxygen, while only 1128 l of oxygen would be absorbed at 141 cc of O_2 per liter of blood. Laurie (1933, p. 374) estimates that the lung capacity of this whale would be 3050 l of air, which would furnish 610 l of oxygen and 2440 l of nitrogen. Krogh (1934, p. 636), however, calculates that the lung capacity of the same whale is 14,000 l of air, which would give 2800 l of oxygen and 11,200 l of nitrogen.

Ommanney (1932, p. 359) has suggested the possibility that the vascular fatty masses around the cervical and anterior dorsal vertebræ may be capable of storing oxygen. Oxygen is quite soluble in fats and the oxygen absorption capacity of whale oil is nearly half that of blood. Some mechanism would be required for a speedy transfer of oxygen to the fatty masses around the retia mirabilia during respiration and for its subsequent liberation into the blood during the period of submergence. Laurie (1933, p. 397) believes that the function of oxygen storage for the fat of the retia mirabilia is improbable since the whale blood has a low hemoglobin content and small oxygen capacity. Calculations based on the actual weights of parts of the carcass indicate that the amount of blood present in the body of a whale is proportionately greater than in land mammals. According to Laurie (1933, p. 373) the volume of blood in a blue whale amounts to 6.6 per cent of the total body weight, and the hemoglobin content (Laurie, 1933, p. 396) of whale blood (9.62 per cent) is less than that of a human (13.8 per cent). The smaller capacity for storage of oxygen in the whale would be disadvantageous if there was a tendency to oxygen shortage during submergence.

The investigation of the vascular system of a fetal finback undertaken by Dr. Robert Walmsley of the Department of Anatomy of the University of Edinburgh while holding a fellowship granted by the Rockefeller Foundation (1935–1936) for research at the Department of Embryology of the Carnegie Institution of Washington furnishes some additional information. Walmsley (1938, pp. 170, 171, 172) concluded that the possibility of oxygen shortage is not a determining factor in the structure and arrangement of the vascular system, that the peculiarities of this system are adaptations to the possible effects of differences in pressure, and that shortage of oxygen does not occur since the quantity of air inhaled before diving is sufficient for the needs of a full submergence. The presence and distribution of the retia mirabilia in the thorax are interpreted as being largely dependent on the diminution of lung volume during submergence. When the hydrostatic pressure is increased the retia in the non-collapsible part of the thoracic cavity become engorged with blood. These retia constitute the efferent supply of the vascular masses which are distributed to the central nervous system and thus insure a more certain and more continuous blood supply during deep and prolonged submergence. According to this explanation, the lungs expand in direct proportion to the decrease in hydrostatic pressure as the whale ascends to the surface and consequently the retia are drained of much of their blood. The generalized condition of the coronary and cerebral arteries and particularly the breaking up of a single trunk into collateral vessels is generally accepted according to Walmsley (1938, p. 173) as indicating that the blood pressure of the whale is low and that low blood pressure is to be associated with a relatively large amount of blood.

Respiratory Demands and Limitations in Relation to Oxygen Consumption

There appears to be no likelihood that the physiology of living whalebone whales will ever be studied by direct observation. Obviously it will be impossible either to capture or to confine one of the larger whales so that metabolism studies can be made. For comparative studies, the ideal conditions for measurement of the minimum basal metabolism are (1) complete muscular repose, (2) cessation of all active digestion, and (3) a temperature near 62.4°F. (28°C.). It is possible that some corroboratory data may be obtained from one of the smaller porpoises. In 1928 following preliminary conversations between Dr. Merriam, Dr. Francis G. Benedict, and myself, plans were made for an attempted investigation of the basal metabolism of a porpoise; but the closing of the bottlenose porpoise fishery at Hatteras, North Carolina, and the lack of facilities elsewhere for the capture of live porpoises made it impossible to carry them out.

Although no experimental data are available on the metabolic rate, a number of calculations of the intensity of normal metabolism based on Rubner's law of surface area (1883, p. 535) and the 27.18-m (89-foot) blue whale measured and weighed at Stromness, South Georgia (Laurie, 1933, p. 404), have been made. By employing the method suggested by Guldberg (1907), Laurie (1933, p. 367) calculated that the approximate surface area of this blue whale would be 275 sq. m. On the assumption that the surface law is valid for whales and that a warm-blooded mammal normally requires a daily expenditure of about 1000 calories per square meter of body surface to carry out the motor processes essential to life, Laurie arrives at a figure of 275,000 calories as the resting metabolism of a 122,004-kg whale, or 2.25 calories per kilogram of body weight per day. Since this calculated metabolic rate is 14.6 times less than that of a man weighing 70 kg and having an oxygen consumption of 4.28 cc per minute per kilogram of body weight, Laurie estimates that this whale would require 35.75 l of oxygen per minute. If one employs the formula ($S = 10 \times W^{\frac{2}{3}}$) suggested by Benedict (1936, p. 285), the surface area of this 122,004-kg whale would be 246 sq. m, and accordingly the resting metabolism would be calculated at 2.016 calories per day per kilogram of body weight, or 16.3 times less than that of the 70-kg man, and thus require 31.96 l of oxygen per minute. Both of these calculations are somewhat higher than the one given by Barcroft (1934, p. 121), who assumes that 20 l of oxygen per minute would be required for basal metabolism. This calculated oxygen consumption for a blue whale of 122,004 kg weight is, as has been pointed out by Benedict (1936, p. 289), but 3 times that (7 l per minute) of an elephant (3672 kg) one-twenty-fifth of that weight. Muscular exertion will greatly increase this oxygen consumption. Calculations of oxygen consumption by a whale while swimming have been published by Krogh (1934, p. 636) and by Barcroft (1934, pp. 122–126).

It would be unprofitable to consider any further these calculations as to the vital capacity of a whale of this size, since experimental work does not appear to substantiate Rubner's body surface law that heat production per square meter of body surface is constant for all mammals. Although the elephant is the largest living land mammal, laboratory tests by Benedict (1936) demonstrate that it has the lowest heat production (13 calories) per kilogram of body weight and the highest heat production (2060 calories) per square meter of body surface of all mammals thus far studied.

The possibility of some control over the metabolic rate, which is alluded to under thermal adjustments, has been discussed with Dr. Oscar Riddle of the Carnegie Institution's Station for Experimental Evolution and with Dr. E. M. K. Geiling of the Department of Pharmacology of the University of Chicago. The results of recent investigations on internal secretions of pituitary glands seem to offer a promising approach to this problem, although it may be questioned whether we now possess the facts necessary to an understanding of the precise mechanism.

The cetacean pituitary gland consists of a large flattened anterior lobe and a smaller neural lobe, but no pars intermedia or residual lumen is present (Wislocki and Geiling, 1936). Pial and dural arteries furnish a rich supply of blood to these glands. The potent pressor component pitressin, which is formed in the neural lobe, has been shown by Geiling's experiments to effect after release a reduction of oxygen consumption. Geiling (1935, pp. 133–134) suggests that this secretion may be one of the factors which enable the whale to remain submerged for considerable periods of time.

Some mechanism to effect the release of this hormone when the animal is submerged is therefore necessary unless the whale possesses some voluntary control over such secretion. One of the hormones that might be expected to assist in effecting such a release is intermedin, generally known as the melanophore principle. Intermedin has been found in the anterior lobe of the whale's pituitary (Geiling, 1935, p. 128). The possibility that there is some sort of relationship between intermedin and the end organs of nerves in the skin is suggested by Ferguson's discovery (Smith, Burr, and Ferguson, 1935, p. 409) that this hormone can be recovered from human melanotic tumors presumed to be neurogenetic in origin. On the upper and lower jaws of cetaceans there are a varying number of peculiarly modified sinus hairs, each of which is supplied with several hundred myelinated nerve fibers from infraorbital and mental nerves, and these fibers connect with numerous tactile Pacinian corpuscles. These structures may afford a mechanism for the appreciation of slight changes in water pressure (Huber, 1934, p. 122) which might either effect the release of intermedin or be communicated directly to the neural lobe of the pituitary. Assuming that water pressure so acts as either to effect a release of intermedin at these specialized nerve endings or to effect direct communication between the Pacinian corpuscles and the neural lobe of the pituitary, thus bringing about a secretion of the fraction pitressin, one may venture an outline of a mechanism by which the whale overdoses himself with this hormone during submergence and ceases this secretion when at the surface. As the whale descends in the water the hydrostatic pressure, secretion, and release go hand in hand, and decreased oxygen consumption ensues as in Geiling's experiments. As the whale rises to the surface the pressure diminishes and the output of pitressin, the pressor component of pituitrin, is either stopped or reduced.

Granting that the views that have been expressed are largely speculative in the light of our present knowledge of the cyclic adjustments in the respiratory mechanism of whales, it is possible that the physiological set-up during a deep submergence may be conditioned by a low basal metabolism. It is possible that owing to the relatively small surface as compared to the weight, the whale's metabolism is so low that it can hold its breath without discomfort for unusually long intervals. Thus the deficit of oxygen per kilogram of body weight, if the metabolism is low, may not exceed the postulated limit, even though a corresponding amount of lactic acid should form in the tissues.

Tolerance to Excess Amounts of Carbon Dioxide

If the physiology of whales is in any way comparable with that of land mammals, the accumulation of carbon dioxide in the blood and body fluids and its diffusion in the alveolar air of the lungs during a prolonged deep submergence will be considerable. Changes in the tension of carbon dioxide gas in the arterial blood are normally proportional to changes in the partial pressure of this gas in the respiratory alveoli. Since the respiratory center of most land mammals is very sensitive to any rise in the tension of the carbon dioxide in the blood that supplies it, an increase in the partial pressure of this gas in the lungs would be expected to be disadvantageous to any deep-diving mammal. It is possible, however, that the respiratory center of whales may be adapted to tolerate a different range of carbon dioxide tensions than exists in land mammals. Laurie (1933, p. 372) is of the opinion that whales have a less delicate mechanism for control of carbon dioxide tension in the blood than land mammals.

Nothing is known in regard to the possible physiological function of such specialized fat organs as the huge spermaceti "case" which imparts the characteristic shape to the head of a sperm whale, the adipose cushion or "melon" located in front of the blowhole on the head of the porpoises and beaked whales, the oil-filled sinuses at the hinder end of each lower jaw of a porpoise, or the intricate and extensive system of intercommunicating oil ducts and sinuses in the spongy subcutaneous and intermuscular tissue of a narwhal. The spermaceti "case" on the male sperm whale's head contains from 10 to 12 barrels of liquid oil. It is not likely that such specializations as these are wholly fortuitous.

Biochemical evidence indicates that the natural nutritive fats of animals eaten as food by whales are resorbed in part unchanged. Klem (1935, p. 50) concludes that the resorbed fat of whales is modified in its composition, but nevertheless must be considered as an admixture of fats of some other kind of animal or plant with fats derived from carbohydrates. He has calculated (1935, p. 95) that a minimum of 50 per cent of the oil of a blue whale is formed by absorption of nutritive fat. The comparatively high content of vitamin A in reddish mysticete whale oil appears to be attributable in part to the presence of the carotenoid pigment astacin, which is derived from krill (*Euphausia superba*) taken as food (Drummond and MacWalter, 1935, p. 106; Collin, Drummond, *et al.*, 1934, p. 207). Krill in turn feed on plankton including the minute plants (diatoms) which may constitute the ultimate source of some at least of these complex carotenoids. Melville (1926, p. 414) likens the aroma given off by the spermaceti oil from a freshly opened spermaceti "case" to the smell of spring violets. Dr. H. A. Spoehr of the Division of Plant Biology, Carnegie Institution of Washington, called my attention to the possibility that this violet odor may result from the oxidation or breakdown of complex unsaturated hydrocarbons, such as carotene. The source of the odor of violets possessed by partially oxidized carotene is attributed to the presence of β-ionone.

There is some evidence to show that slight differences in the complex chemical compounds of these carotenoids may be of considerable significance in the functioning of these substances, when they occur in a combined state with fat. So far as is known, no attempt has been made to isolate these compounds from whale fats and study their properties. In view of the part these compounds play in the physiology of plants (diatoms included), the possibility is presented that they, through the intervention of some reaction of an enzymatic nature or in response to increased hydrostatic pressure and the colder temperature of the surrounding water during a deep submergence, may subserve some such function as the absorption of excess quantities of carbon dioxide so as not to overtax the tolerance of the respiratory center during deep submergence when otherwise an abnormal rise in the partial pressure of this gas in the lungs and of its tension in the blood might be expected.

LIABILITY OF WHALES TO THE EFFECTS OF EXCESS AMOUNTS OF NITROGEN

During a prolonged deep submergence, it might be expected that the blood passing through the lungs would become supersaturated with nitrogen. Laurie (1933), however, found on analysis of arterial blood samples from more than 80 whales that there was actually less nitrogen in solution than should have been dissolved normally in accordance with established physical laws, notwithstanding the high pressures that had been encountered by these whales during the dives after being harpooned. These pressures should have caused the solution of greater volumes of gas than under normal atmospheric conditions. When control experiments were performed to ascertain the solubility of air nitrogen in whale's blood, it was found that more than the normal amount of nitrogen, that is 1.20 vol. per cent, had been taken up. The nitrogen capacity of one sample was found to be 3.20 vol. per cent. Laurie (1933, p. 389) is convinced that supersaturation is bound to occur if the whale's dive is sufficiently deep, and that the whale has a special mechanism for disposing of the dissolved nitrogen in the blood, since whale blood when examined after the whale emerges from a dive is not supersaturated but is actually subsaturated.

A human diver encounters the danger of effervescence of the dissolved nitrogen in the tissues, if he ascends so rapidly that the pressure is reduced more quickly than the blood can transport the released nitrogen to the lungs. These nitrogen bubbles obstruct circulation and produce diver's sickness. No additional air, however, is pumped to the whale while submerged, as is the case with the human diver or caisson worker, and the animal utilizes the supply of air that fills the lungs at the time of submergence. The available nitrogen is thus limited to that contained in the air in the lungs.

Granting that the pressure within the thorax including that of air within the whale's lungs must be equalized with the hydrostatic pressure on the outside, at a depth of 100 m the total pressure would be about 11 atmospheres. Captain Damant (1934) therefore asserts that the whale's lungs would be compressed until an average alveolus would have only one-tenth of its normal volume. This shrinking of the alveoli must decrease the surface available for diffusion; the epithelium of each individual alveolus must become thicker and thus further retard diffusion. Under such conditions the entrance of excess nitrogen into the blood would be obstructed and conversely the rapid discharge of nitrogen would be facilitated after decompression when the animal is breathing at the surface. Ommanney (1935) likewise believes that the high contractibility of the alveolar walls of the lungs and the slower circulation rate reduces a whale's susceptibility to diver's sickness.

Since a whale's nitrogen capacity has been calculated at about 2 per cent of the animal's volume per atmosphere pressure, Irving (1935, p. 560) contends that four atmospheres extra hydrostatic pressure, that is a submergence of 40 meters, will cause the solution of all lung gases, provided total collapse of the lung ensues. But if all the nitrogen is dissolved into the tissues, the amount is still only three times as great as any experienced human diver can safely eliminate. During decompression while rising to the surface, the circulation may be accelerated by the stimulation of the whale's large oxygen debt. Irving believes that the whale's circulation will be three times as effective as that of a human diver at the time of emergence from a deep dive. He therefore concludes that in view of the limited amount of nitrogen available for solution in the blood and tissues and the favorable conditions of the circulation the whale on returning to the surface after a deep submergence should not be sensitive to the conditions which cause diver's sickness. Other investigators (Walmsley 1938, p. 173; Laurie, 1933, p. 392) hold the view that the architecture of the whale's vascular system indicates a relatively low blood pressure and a sluggish circulation. After decompression the respiratory epithelium presents its normal surface for subserving respiratory functions. Furthermore, the unusual quantity of elastic tissue and intrinsic musculature in the pleura and the bronchioles of the lungs and the oblique position of the diaphragm constitute a powerful mechanism for emptying the lungs in a minimum of time. In view of the rapid ascent of the whale from a deep submergence, Sir Leonard Hill has expressed the opinion (1935) that the quantity of nitrogen in the air contained in the lungs is not sufficient to produce bubbles of nitrogen in the blood during decompression, an entirely reasonable assumption.

Modifications of the Skull and Skeleton

The foregoing discussion has been limited to conditions that affect most obvious vital processes of a cetacean. There are, however, many equally interesting and complicated alterations in other portions of the whale's anatomy. The adaptation of structure to function is likewise illustrated by the manifold alterations that have been effected in the skulls and skeletons of cetaceans during the course of geologic time. One of the objects of these palæobiological studies has been to discover how the characteristic skeletal and cranial peculiarities of each of the several families were acquired, how the members of these families have arrived at their present form, and to what extent functional adaptations have modified osseous structures.

The exigencies of an aquatic mode of living are such that the members of the several families of the three suborders of Cetacea assumed a similar outward appearance in conformity with streamline requirements, a mechanically induced convergence which is likewise manifested in many other unrelated groups of aquatic vertebrates; but there is no palæontological evidence to support any assumption of a remodeling of originally unlike parts to a similar adaptive use. The directions in which the members of the three cetacean suborders, collectively and individually, subsequently developed were predetermined by peculiarities in existence in the skulls of their progenitors before the tendency toward the foreshortening of the braincase and the shifting of the bones to new positions had been made manifest. The old Tertiary archæocetes, however, had already acquired the fundamental adaptive alterations which were subsequently perfected to varying degrees and in different directions by later odontocetes and mysticetes. With the exception of the archæocetes, the skulls of all known cetaceans show in some degree the effects of telescoping; that is, the braincase or portion of skull behind the rostrum has been shortened, by the slipping of one bone over another and by interdigitation.

An examination of the skulls and skeletons of these extinct cetaceans shows clearly that alterations of different degrees and in different directions have been effected in all parts of the skeleton and that in Recent cetaceans comparatively little of the original construction has been retained. Structural alterations, that have been effected in the skull either by telescoping or by the shifting to new positions and relations of the component bones, have brought into existence a wide variety of cetacean types. The palæontologist, while preparing descriptions of these various types, generally attempts to interpret what these structural modifications signify as regards possible relationships with other known forms, and also to what extent adaptive alterations in the anatomy are reflected in skeletal structures. Nevertheless the significance of such homological resemblances and differences can rarely be appraised until the general adaptive trend and possible change of function of each part is fully understood and critically analyzed. Inasmuch as some of the evidence relating to the geological record of the Cetacea has already been reviewed elsewhere (Kellogg, 1928; Kellogg, 1936), this phase of the subject will not be elaborated further here.

There is much that we do not as yet know about the geological history of the Cetacea and the adjustments that have been made so that the vital physiological processes will not be impaired. Nevertheless, one cannot fail to be impressed by the intricacy of the adaptive alterations that have so shaped countless anatomical details that each now seems to subserve a specific function. The available biological evidence is, however, not sufficient to warrant the assertion that any of the interpretations that have been given can or cannot be final. Our conceptions of the functional anatomy of the Cetacea are being constantly revised and it is not unlikely that many of the interpretations that appear reasonable at the present moment will be proved untenable by the new facts that the future may be expected to bring to light.

LITERATURE CITED

Barcroft, Joseph, 1934. Features in the Architecture of Physiological Function. New York, The Macmillan Company, x + 368 pp., 106 figs.

Bartholin, Thomas, 1654. Historiarum anatomicarum rariorum. Centuriae i et ii, historiae xxv, Amstelodami, J. Henry, 326 pp., 9 pls.

Benedict, Francis G., 1936. The Physiology of the Elephant. Carnegie Inst. Wash. Pub. No. 474, vii + 302 pp., 13 figs., 8 pls.

Collin, G., J. C. Drummond, T. P. Hilditch, and E. R. Gunther, 1934. Observations on the Fatty Constituents of Marine Plankton, II: General Characters of the Plankton Oils, *Jour. Exper. Biol.* (London), vol. 11, no. 2, pp. 198–209, April, 1934.

Damant, G. C. C., 1934. Physiology of Deep Diving in the Whale, *Nature* (London), vol. 133, no. 3371, p. 874, June 9, 1934.

Drummond, J. C., and R. J. MacWalter, 1935. Provitamin A in the Food of Whales, *Jour. Exper. Biol.* (London), vol. 12, no. 2, pp. 105–107, April 1935.

Geiling, E. M. K., 1935. The Hypophysis cerebri of the Finback (Balænoptera physalus) and Sperm (Physeter megalocephalus) Whale, *Bull. Johns Hopkins Hospital*, vol. 57, no. 3, pp. 123–142, 2 pls., September 1935.

Guldberg, G., 1907. Über das Verfahren bei Berechnung des Rauminhaltes und Gewichtes der grossen Waltiere, *Forhandlinger Videnskabs-Selskabs* (Christiania), no. 3, pp. 1–12.

Haynes, F., and Alec H. Laurie, 1937. On the Histological Structure of Cetacean Lungs, *Discovery Reports* (Cambridge, England), vol. 17, pp. 1–6, 1 pl., July 1937.

HILL, LEONARD, 1935. [Do Whales Descend to Great Depths?], *Nature* (London), vol. 135, no. 3417, p. 657, April 27, 1935.

HOWELL, A. BRAZIER, 1927. Contribution to the Anatomy of the Chinese Finless Porpoise Neomeris phocænoides, *Proc. U. S. Nat. Mus.*, vol. 70, art. 13, pp. 1–43, 14 figs., 1 pl., March 11, 1927.

——— 1930. Aquatic Mammals: Their Adaptation to Life in the Water. Springfield and Baltimore, Charles C. Thomas, xii + 338 pp., 53 figs., 1 pl.

HUBER, ERNST, 1934. Anatomical Notes on Pinnipedia and Cetacea. Carnegie Inst. Wash. Pub. no. 447, pp. 105–136, 12 figs., January 20, 1934.

HUNTER, JOHN, 1787. Observations on the Structure and Oeconomy of Whales, *Philos. Trans, Roy. Soc. London*, vol. 77, pp. 371–450, pls. 16–22.

IRVING, LAURENCE, 1935. The Protection of Whales from the Danger of Caisson Disease, *Science*, n. s., vol. 81, no. 2110, pp. 560–561, June 7, 1935.

KELLOGG, REMINGTON, 1924. Tertiary Pelagic Mammals of Eastern North America, *Bull. Geol. Soc. Amer.*, vol. 35, pp. 755–766, December 1924.

——— 1928. The History of Whales—Their Adaptation to Life in the Water, *Quart. Rev. Biol.*, vol. 3, no. 1, pp. 29–76, figs. 1–11, March 1928; no. 2, pp. 174–208, figs. 12–24, June 1928.

——— 1929. What Is Known of the Migrations of Some of the Whalebone Whales, *Smithsonian Ann. Rept. for 1928*, pp. 467–494, 6 figs., November 18, 1929.

——— 1936. A Review of the Archæoceti. Carnegie Inst. Wash. Pub. no. 482, xv + 366 pp., 88 figs., 37 pls., December 14, 1936.

KLEM, ALF, 1935. Studies in the Biochemistry of Whale Oils. *Hvalrådets Skrifter, Sci. Res. Marine Biol. Research* (Oslo), no. 11, pp. 49–108.

KROGH, AUGUST, 1934. Physiology of the Blue Whale, *Nature* (London), vol. 133, no. 3365, pp. 635–637, April 28, 1934.

——— 1934*a*. [Comments on Article by Captain G. C. C. Damant], *Nature* (London), vol. 133, no. 3371, p. 874, June 9, 1934.

LANGWORTHY, ORTHELLO R., 1932. A Description of the Central Nervous System of the Porpoise (Tursiops truncatus), *Jour. Compar. Neurol.*, vol. 54, no. 2, pp. 437–488, 8 figs, 7 pls., April 15, 1932.

LAURIE, ALEC H., 1933. Some Aspects of Respiration in Blue and Fin Whales, *Discovery Reports* (Cambridge, England), vol. 7, pp. 363–406, 4 figs., pl. 15, December 1933.

——— 1935. Physiology of Whales, *Nature* (London), vol. 135, no. 3420, p. 823, May 18, 1935.

MEIGS, CHARLES D., 1849. Observations on the Reproductive Organs and on the Fetus of the Delphinus Nesarnak, *Jour. Acad. Nat. Sci. Philadelphia*, n. s., vol. 1, pt. 3, pp. 267–273, pls. 35–36, August 7, 1849.

MELVILLE, HERMAN, 1926. Moby Dick. New York, The Modern Library, xxi + 565 pp.

MILLER, GERRIT S., JR., 1923. The Telescoping of the Cetacean Skull, *Smithsonian Misc. Coll.*, vol. 76, no. 5, pp. 1–70, 8 pls., August 31, 1923.

MYERS, ROLLIN G., 1920. A Chemical Study of Whale Blood, *Jour. Biol. Chem.* (Baltimore), vol. 41, no. 1, pp. 137–143, January 1920.

Ommanney, F. D., 1932. The Vascular Networks (Retia mirabilia) of the Fin Whale (Balænoptera physalus), *Discovery Reports* (Cambridge, England), vol. 5, pp. 327–362, 10 figs., September 1932.

——— 1932*a*. The Urino-genital System of the Fin Whale (Balænoptera physalus), Appendix: The Dimensions and Growth of the Kidney of Blue and Fin Whales, *Discovery Reports* (Cambridge, England), vol. 5, pp. 363–466, 39 figs., pls. 2–3, September 1932.

——— 1935. Do Whales Descend to Great Depths? *Nature* (London), vol. 135, no. 3411, pp. 429–430, March 18, 1935.

Peters, Nicolaus, 1938. Über die Ernährung der südlichen Bartenwale, *Natur und Volk* (Frankfurt a. M.), vol. 68, no. 3, pp. 124–130, 8 figs., March 1, 1938.

Portier, P., 1910. Pression osmotique des liquides des oiseaux et mammifères marins, *Jour. Physiol. et Pathol. Gén.* (Paris), vol. 12, pp. 202–208.

Ray, John, 1671. An Account of the Dissection of a Porpess. *Philos. Trans. Roy. Soc. London*, vol. 6, no. 76, pp. 2274–2279.

Ries, Ferd August, and Orthello R. Langworthy, 1937. A Study of the Surface Structure of the Brain of the Whale (Balænoptera physalus and Physeter catodon), *Jour. Compar. Neurol.*, vol. 68, no. 1, pp. 1–47, 4 pls., December 15, 1937.

Rubner, Max, 1883. Über den Einfluss der Körpergrösse auf Stoff- und Kraftwechsel, *Zeitschr. Biol.* (München u. Leipzig), (n. s., vol. 1), vol. 19, pp. 535–562.

Smith, G. M., H. S. Burr, and R. S. Ferguson, 1935. A Study of the Effects of Intermedin and Injury of the Hypophysis on Traumatic Corial Melanophores in Goldfishes, *Endocrinology*, vol. 19, no. 4, pp. 409–412, July–August 1935.

Sudzuki, M., 1924. Untersuchungen über Cetacea, VIII: Über das Blut, *Tohoku Jour. Exper. Med.* (Sendai), vol. 5, nos. 4–5, pp. 419–427.

Walmsley, Robert, 1938. Some Observations on the Vascular System of a Female Fetal Finback. Carnegie Inst. Wash. Pub. no. 496 (Contrib. Embryol. No. 164), pp. 107–178, 27 figs., pls. 1–5, tables, May 31, 1938.

White, W. Hale, 1891. A Theory to Explain the Evolution of Warm-blooded Vertebrates, *Jour. Anat. and Physiol.* (London), vol. 25 (n. s., vol. 5), pp. 374–385.

Wislocki, George B., 1929. On the Structure of the Lungs of Porpoises (Tursiops truncatus), *Amer. Jour. Anat.*, vol. 44, no. 1, pp. 47–77, 3 pls., July 15, 1929.

——— 1933. On the Placentation of the Harbor Porpoise (Phocæna phocœna [Linnæus]), *Biol. Bull.*, vol. 65, no. 1, pp. 80–98, 5 pls., August 1933.

——— and E. M. K. Geiling, 1936. The Anatomy of the Hypophysis of Whales, *Anat. Rec.*, vol. 66, no. 1, pp. 17–41, 4 figs., 3 pls., August 25, 1936.

STRUCTURES IN MODERN SEDIMENTS AID IN INTERPRETING ANCIENT ROCKS

EDWIN D. McKEE

Park Naturalist, Grand Canyon National Park
Research Associate, The Carnegie Institution of Washington

INTRODUCTION

THE struggle that represented the development of geology into a recognized science was an epic of man's perversity, featuring especially his frequent willingness to be guided in thought by preconceived ideas, by local viewpoints, and by faith in the supernatural. The progress of geology was most hindered near its inception, because students "took not their material from Nature, but spun it out of themselves." The history of this science, furthermore, shows many early failures in the interpretation of commonplace features, largely because geologists shirked the labor of making observations and tried to supply its place by speculation based on the imagination.

James Hutton has aptly been called the founder of geology since it was through his successful efforts late in the eighteenth century that students first realized the necessity of understanding earth processes as they operate today before attempting to unravel the history of the earth. Early in the nineteenth century the teachings of Hutton were revived and illustrated by Sir Charles Lyell, and they are now universally recognized as the basis of all development in the field of general geology.

Since the time of Lyell, progress in geological science has been steady and rapid. This advance has been possible, however, only because of the vast number of facts that have been recorded throughout the world as a result of detailed observations of natural processes and of their results. These represent the groundwork on which the accuracy of more speculative endeavors depends. Despite all the refinement, specialization, and exactness found in certain aspects of geology today, it is apparent that certain phases of it still are very poorly understood, largely because even now few data are available on the operation of the particular processes which they represent.

To the student of sedimentary rocks, one of the best illustrations of present-day limitations due to lack of recorded observations of the processes is in the field of original structures, especially those classified as beds or laminæ. Such structures are frequently the most apparent features with which a geologist has to deal when studying sedimentary rocks, yet their significance, for the most part, must remain obscure or be interpreted on the basis of pure speculation simply because a record of modern counterparts has not been made.

The task of recording data on original structures in sediments is not altogether untouched. Nearly fifty years ago one of America's greatest geologists, G. K. Gilbert, critically examined and described the bedding of small deltas formed on the margins of old Lake Bonneville, at the same time clearly demonstrating the characteristics of small deltas of that particular type (4; see "Literature Cited" at end of paper). Recent studies by Thompson of beaches on the California coast and of the structures in one variety of sand dune have pointed the way toward a systematic classification and analysis of bedding types (12). A few other studies have been made and numerous scattered observations recorded from time to time, yet the fact remains that today most bedding and lamination structures prove more puzzling than helpful to the geologist who endeavors to interpret the environment of their deposition.

Classification of Environments

In the study of sedimentary rocks, whether for economic or for purely scientific reasons, it is always of value and of fundamental importance to have an understanding of the types of environment represented. In addition to the strictly marine deposits which make up a large portion of all sedimentary formations, there are also many rocks developed as delta, dune, beach, lagoonal, and other deposits. Each of these, in turn, may be subdivided into various types. In the case of delta development, marked differences exist between those of humid and those of arid areas, and between those formed in the proximity of mountains and those far removed from elevated regions. These and many other factors so affect the form, extent, and especially the structure of sedimentary deposits that it is highly important that diagnostic criteria of all types be determined, classified, and compared.

In many deposits recognition of the environment of deposition is made possible, or at least facilitated, by the presence of fossils. In others, knowledge of the physiographic characteristics or of texture may throw light on the origin. A very great amount of detailed information has been recorded on the morphology of deltas, dunes, and other types of deposits, and the quantity of data available from the study of textural characteristics is shown by the fact that more than forty papers dealing with grain shape and twenty on grain size are to be found in a recent bibliography. Original structures in sediments, on the other hand, are at present poorly understood, despite the facts that unlike physiographic form they are usually easily observed and are frequently distinctive in character, and that they definitely represent the environment of deposition rather than that of transportation or source as may be the case with textural features.

Advantages to be gained by dependable methods of classifying sedimentary structures are numerous. In addition to giving obvious assistance in the recognition of general environmental factors, such a classification may elucidate the particular processes involved in deposition of any sediment and it may make possible an estimate of the approximate time involved in a type of deposit. With this in mind the writer has recently been carrying on as a project of the Carnegie Institution of Washington a program of investigation of original structures in the Colorado River delta (8) (9) and in certain sand-dune areas of the region. It is to be hoped that eventually the program can be extended until the structures characteristic of all the principal environments of deposition are known and can be classified and compared. Only then will it be possible to interpret with reasonable assurance the history of many formations now obscure.

THE STUDY OF ORIGINAL STRUCTURES

Studies of original structures as found in ancient rocks have been carried on by many geologists, although for the most part only of specialized cases and usually without detailed or critical analysis. One of the most important contributions of recent years in this field is the work of Dr. S. H. Knight (6), who has clearly shown the desirability of recording and analyzing structures in three dimensions wherever possible. Naturally this principle applies not only to consolidated sediments, but also to those accumulating today. The latter, fortunately, are usually well adapted to such study.

In modern deposits of deltas, beaches, and other subaerial environments, natural sections exposing faces which illustrate three dimensions of a structure are often not available for study. In such places, however, it is usually possible to cut the desired sections by careful use of a large knife. An exception to this is the average sand dune, which is too dry to stand with clean, vertical walls when sectioned. Only in a few favorable localities such as that of the gypsum dunes near Alamogordo, New Mexico, where natural cohesive material exists as a bond in the sediment, or along certain coastal areas where moisture serves the same purpose, can sections suitable for dune study be obtained. It has been demonstrated, however, that through artificial dampening of dunes, comparable conditions can be obtained elsewhere (1).

Modern marine deposits have never been examined in regard to structures, so far as the writer knows, although many geologists have advanced suggestions as to what the character of the bedding might be and what types of cross-lamination would be formed where marine currents or wave action are strongly developed. Obviously the difficulties in recording structures under water are numerous, but it is conceivable that vertical sections might be exposed in shallow-water sediments by inserting a vertical glass sheet and then removing the deposits on one side. Problems of adequate lighting and means of recording the structures would then need to be overcome. A more easily developed method of recording such data is that of taking core samples. This has the disadvantage of making accessible only a small exposure in any one core, but it has the advantages of greater ease in obtaining material and in making examinations and it offers the possibility of securing specimens from fairly great depths.

The adaptability of different sediments for lamination analysis varies tremendously. Where adjacent laminæ are characterized by marked contrast in grain size, composition, or both, original structures are readily apparent on every exposed face. Frequently, however, laminæ are not of these types. They develop in sediments of essentially homogeneous character, apparently as a result of parting planes formed where thin films or crusts of salt have been deposited, or along zones of less compaction than normal. In such instances only a weathered surface will show details of lamination. Thus in new or artificially developed sections, etching of the faces by blowing must be resorted to in order to bring out the structure patterns.

As a result of preliminary work in this field which has been undertaken on the delta of the Colorado River in Mexico, the writer has found that cross-lamination may logically be classified into two major types: (1) that involving deposition with nearly contemporaneous partial erosion and (2) that involving deposition alone. Furthermore, it is possible to demonstrate to which group a particular structure belongs provided a horizontal section through it can be obtained. In the case of deposition without erosion, laminæ appear like contours on a topographic map, sometimes converging and meeting but never crossing one another. In the other type, lines representing erosion surfaces naturally cut across those of deposition.

Further classification of cross-lamination should probably be based on such features as the general size, shape, and dip of depositional planes, and, where involved, the character and slope of successive erosion surfaces. It does not seem worth while to attempt any such classification at present, however, because knowledge of the many types—where each occurs and how it is formed—is still very incomplete. As definite information accumulates on the forms characteristic of each particular environment, a much needed classification can be worked out and this will greatly assist the geologist in making precise descriptions of sedimentary rocks and in interpreting them.

Thompson in his studies on modern beaches in California (12) has demonstrated, through the use of measuring stakes, the feasibility of obtaining definite data on the time represented by deposits of a given type. This kind of investigation is obviously very important in the field of sedimentation and measurements of time should be carried on wherever possible in connection with studies of bedding and lamination.

Another feature of cross-laminated sediments which has been given virtually no attention in modern deposits is the dominant direction of dip of laminæ as found over wide areas. When sufficient data have been recorded to show the variability or persistence of this dip in specific environments, more accurate interpretations of many ancient sedimentary rocks will result. The possibilities in this type of study have already been clearly illustrated by the systematic plotting of structures in certain æolian sandstones in England which have been studied by Dr. F. W. Shotton (11), and by somewhat similar statistical studies made in America by Dr. Parry Reiche (10).

Inferences Regarding the Origin of the Coconino Sandstone Based on Study of Modern Structures

The Coconino sandstone, which forms a prominent white cliff in the upper walls of Grand Canyon, is distinctive because of the uniformity and purity of its sand grains and because of the large scale on which it is cross-laminated. It is one of the most conspicuous formations in the region, and one which probably has given rise to more controversies regarding the nature of its origin than has any other.

In 1933 a detailed study of the Coconino sandstone was completed by the writer (7), and at that time the conclusion was reached that this stratum was formed almost entirely as the result of sand having been piled up into dunes by the wind. Although several different types of evidence led to this conclusion, the most apparent and widespread feature of the formation—its cross-lamination—failed to lend any evidence either pro or con, for no data on the structure of modern dunes were then available for comparison.

Subsequently to this work on the Coconino and despite evidence based on texture and physiographic form to the contrary, some geologists have remained skeptical of the æolian origin of the formation. The reason has been a concept, developed on theoretical grounds, that all dune structure should show the arc-like lamination pattern that results from truncation of the leeward ends of either lobate or concave dunes. It should be pointed out, however, that even in these types, sections cut in the general direction of wind movement will show steeply dipping laminæ which tend near their bases to become tangent to the underlying surface. Furthermore, in straight-fronted dunes, since the laminæ do not develop the form of plunging anticlines or synclines, no sections show arc-like patterns. Laminæ in them are essentially flat sheets which dip leeward with bases that curve and become tangent to the floor of deposition. In all these types of sand dunes, moreover, the upper surfaces of each set of laminæ may develop into any of several forms. Depending on the type of dune, they will range in shape from narrow, chute-like valleys to broad, flat surfaces either with or without deposition of windward-slope sediments on top. In angle of dip, which is always low, they will vary according to the ratio between the supply of sediment available for deposition and the wind velocity, since these factors control the amount of erosion or of deposition on the windward side of the dune. Thus, depending on type of dune and direction of sectioning, a tremendous number of variations may be expected in lamination, in truncating surfaces, and in the relations of the one to the other.

On the south shore of Lake Mead near Iceberg Canyon are some excellent examples of modern straight-fronted dunes which, by virtue of recent undercutting due to rising waters and because of some undetected but effective bond between sand grains, develop cliff faces in which are demonstrated the lamination patterns of longitudinal sections. Similarity between these and the original structures typical of most of the Coconino is marked (see plate 1). In both deposits lee-side laminæ dip with high angle, are frequently very long, and are tangential at base. They are variable in direction and angle of dip not only in different groups but also within any one set. Horizontal laminæ are absent but some erosion planes are nearly horizontal and most of them extend for many feet with relatively flat surfaces that incline with dips of only a few degrees toward the direction of wind. In brief, the unusual features of the typical cross-lamination in the Coconino sandstone are adequately explained by comparison with and analysis of modern straight-fronted dunes such as those found along Lake Mead.

Examination of the great sand areas of the Sahara desert today suggests that the long, ridge-like and relatively straight-fronted dune, rather than the small crescentic or barkhan dune, should be expected to predominate in the development of an extensive deposit such as the Coconino sandstone. E. F. Gautier, professor in the University of Algiers, has stated in his classic description of the Sahara desert (3) that "the typical barkhan is very rare" in this desert, and that of the several dune forms which are definite features of the topography, only the sword-like *sifs* "might bear some distant relation to the barkhan." It is clear then that the absence of arc-like lamination structures, such as are developed in the small crescentic dunes that are so common locally in western America and in the deserts of Turkestan and Mongolia, and which are prominent features of certain ancient, supposedly wind-deposited formations in northern Arizona, is not inconsistent with the view that the Coconino sandstone represents an environment comparable to that of parts of the Sahara.[1]

Inferences Regarding the Origin of the Temple Butte Limestone Based on Study of Modern Structures

Rocks of Devonian age in Grand Canyon, designated the Temple Butte limestone, are today more imperfectly known as regards distribution and extent and less understood than are those of any other Paleozoic formation. One reason is their inaccessibility. Since they are overlain by the massive, cliff-forming Redwall limestone, 500 to 800 feet thick, access from above is impossible in a majority of localities, whereas the climb from below is also difficult at best, with numerous cliffs and usually as much as 2000 feet of rocks to ascend. A second factor which tends to keep the Devonian history of the area in obscurity is the unusual character of the rocks of this age—rocks which are not similar to any others in this region, and in which most normal criteria of the environment of deposition are not apparent.

From the vicinity of Bass Trail to the eastern end of Grand Canyon, the Devonian period is represented only by small remnants or "pockets" of sediment of which about fifteen have been recorded to date. West of this area, Devonian rocks form continuous layers that everywhere separate sediments of Cambrian age from those of the Mississippian Redwall limestone above. The "pockets" of eastern Grand Canyon obviously are formed from channels eroded in the surface of the uppermost Cambrian rocks and subsequently filled with sediments. In depth they range up to about 100 feet. Similar pockets of Devonian material are common in western Grand Canyon, locally extending down far below the normal top of the Cambrian, but everywhere overlain by the widespread strata of Devonian age already mentioned. These upper deposits increase in thickness westward from less than 100 feet near Bass Trail to about 1000 feet near the western end of Grand Canyon. Field evidence suggests that this remarkably great increase in a relatively short distance is not to be explained alone by greater sedimentation in the west, but also by increased subsequent erosion eastward.

---

[1] Other sand areas in which crescentic form in dunes is rare are the Libyan desert and northern Indiana. In the former, according to Kadar (5), the longitudinal dune, i.e., ridge parallel to prevailing wind direction, is the dominant type, whereas in Indiana a linear form consisting of a ridge at right angles to the direction of prevailing wind, as described by Cressey (2), makes up much of the dune area.

Three principal types of lithology are recognized in the Temple Butte limestone. One is a dark gray crystalline, massive limestone which weathers to thick beds with olive brown, sandy surfaces and which forms resistant cliffs and ledges. A second is a saccharoidal limestone of distinctive lavender color, structureless or irregularly bedded, and crumbly in most places. The third type is dense, white or pale gray limestone, thin-bedded and usually highly cross-laminated on a small scale. It normally forms slopes or benches which are conspicuous because of their creamy whiteness. Each of these types is present at several different horizons in the thick sections of western Grand Canyon, but in the small Devonian pockets farther east, one or the other of the first two types only is represented.

In the Temple Butte limestone fossils have been found exclusively in the sandy, thick-bedded limestone members. Though very common in parts of these, they are consistently of poor preservation. *Bothriolepis*, a fresh- or brackish-water fish, found at Sapphire Canyon by Dr. L. F. Noble, and numerous gastropods found by the writer in sections farther west strongly suggest that the environment of deposition was one that was neither continental nor marine. It seems probable that this formation represents a littoral or lagunal facies developed contemporaneously with the extensive Devonian coral- and brachiopod-bearing marine deposits of Nevada sections to the west.

The structureless or gnarly-bedded lavender limestones of the Temple Butte are very irregular in both vertical and horizontal distribution and appear to be largely channel-fill deposits. Their structure implies nothing regarding environment, other than instability of depositional factors.

The white, thin-bedded limestones of the Temple Butte apparently offer the best opportunities for obtaining data indicative of the environment represented by this formation. The widespread, persistent cross-lamination is distinctive with its small scale and peculiar form and it no doubt represents very specialized conditions. By three-dimension analysis, the writer has been able to show conclusively that this structure is the result of irregularly superimposed ripple laminæ and is similar in every respect to cross-lamination found prevalent in fine sands over a large part of the upper end of the Colorado River delta today (plate 1).

By the comparison just made, the writer does not wish to imply that the environment of the Temple Butte was comparable to that of the present upper delta of the Colorado; differences in lithology alone prohibit such a suggestion. On the other hand, it seems reasonably certain that the same general conditions of sedimentation, such as the amount of current action and the depth of water, must have prevailed in both environments. Thus, the cross-lamination in the Devonian limestone, when better understood in terms of modern environments, will doubtless prove of great importance in the interpretation of this formation.

Inferences Regarding the Origin of the Supai Formation Based on Study of Modern Structures

The Supai formation of Grand Canyon is a thick series of red sandstones and shales, but contains some limestone units in the lower part which increase in number westward. From a physiographic viewpoint, it is a series of alternating cliffs and slopes located immediately above the Redwall limestone cliff or middle portion of the Canyon.

The deposits which make up the Supai formation have for many years been considered, largely on the basis of land plants and the tracks of vertebrate animals which they contain, to be continental in origin. The exact character of the environment represented by these alternating mud and sand deposits, however, is still poorly understood, and it seems probable that here again original structures such as bedding and cross-lamination will ultimately prove to be of major importance in making interpretations.

Recent studies of bedding in the delta of the Colorado River in Mexico (8) make quite clear the genesis of at least one type of structure found throughout many of the sandstone members of the Supai formation. Reference is made to the long, sloping laminæ, which are straight except near the very base where they curve and become tangent to the floor of deposition, and which occur within beds having flat upper and lower surfaces (plate 1). This pattern seen in many cliff faces is almost certainly the result of sheet-like layers of sand having been deposited in orderly succession, without appreciable change in dip or direction. The individual laminæ are the fore-set beds of local deltas or cones of the type that is common on the middle portions of the present Colorado delta. Here, in times of flood, a major portion of the tractional load is deposited rapidly, the sand sloping down to meet a surface already relatively flat. Because of accompanying beveling of the upper surface, laminæ are usually truncated to a plane which is normally covered in turn by structureless mud deposits dropped in quiet waters after recession of the flood.

Several types of cross-lamination other than the one just described are found in the Supai formation; however, studies in this field have not yet advanced far enough to justify an analysis and a comparison of present and ancient types at this time. It is hoped that eventually exact data on the conditions represented by each of these types will be available, and so permit a more accurate interpretation of conditions during Supai deposition than is now possible.

The Future of Investigations of Original Structures in Modern Deposits

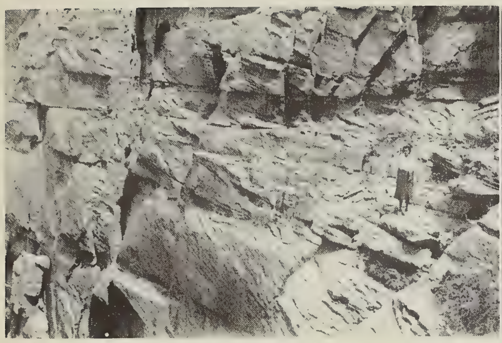
The three illustrations just given of the value of a clear understanding of the origin and nature of various types of bedding and cross-lamination demonstrate the desirability of making further investigations in this field. In some instances, such structures are the only good criteria of the environment represented; in others they substantiate or clarify theories postulated on the basis of fossils, texture, or morphological evidence. In any case, they constitute an important approach to the solutions of various problems.

In order to realize the maximum value of original structures as aids in interpreting past environments, it is necessary that much work be done in recording the various types as found in modern deposits. It is important to recognize not only all the varieties that represent a particular environment, but also all the environments in which may be formed a particular variety. Likewise it is important that a classification be developed by which individual types may be adequately described and designated.

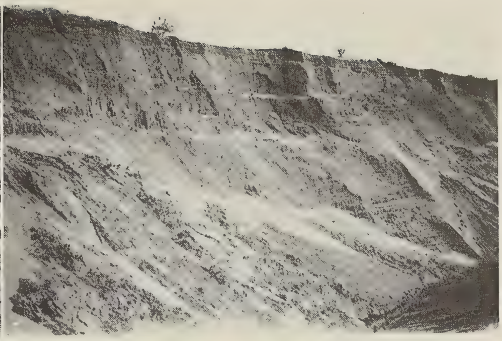
The program of investigation which has been briefly described and discussed in this paper is as yet little advanced, but the field holds such promise for future development of the science that it deserves more attention. The geologist of today will do well to bear in mind the principal stated by G. K. Gilbert many years ago that "the great importance of the chapter on the latest period lies in the fact that it [contains] . . . the key for the decipherment of the records of the earlier."

LITERATURE CITED

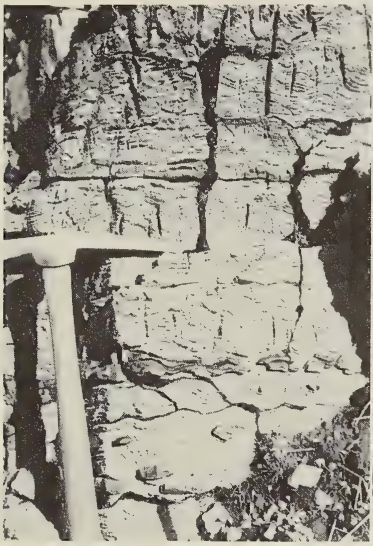
1. Beadnell, H. J. L., The Sand Dunes of the Libyan Desert, *Geog. Jour.*, vol. 35, no. 4, 1910.
2. Cressey, George B., The Indiana Sand Dunes and Shore Line of the Lake Michigan Basin, *Geog. Soc. Chicago*, *Bull. No. 8*, 1928.
3. Gautier, E. F., Sahara—The Great Desert (translation by Dorothy Ford Mayhew), Columbia Univ. Press, pp. 1–264, 1935.
4. Gilbert, G. K., Lake Bonneville, *U. S. Geol. Surv. Monograph 1*, pp. 1–438, 1890.
5. Kadar, Ladislas, A Study of the Sand Sea in the Libyan Desert, *Geog. Jour.*, vol. 83, no. 6, 1934.
6. Knight, S. H., The Fountain and the Casper Formations of the Laramie Basin, *Univ. Wyo. Publ. in Science*, *Geol.*, vol. 1, no. 1, pp. 1–82, 1929.
7. McKee, Edwin D., The Coconino Sandstone—Its History and Origin, Carnegie Inst. Wash. Pub. No. 440, VII, pp. 78–115, 1933.
8. ——— Some Types of Bedding in the Colorado River Delta, *Jour. Geol.* In press.
9. ——— Original Structures in Colorado River Flood Deposits of Grand Canyon, *Jour. Sedimentary Petrol.* In press.
10. Reiche, Parry, New Data on the Coconino Sandstone. Paper presented at 18th Ann. Meeting, Southwestern Div. Amer. Assoc. Adv. Sci., April 24, 1938.
11. Shotton, F. W., The Lower Bunter Sandstones of North Worcestershire and East Shropshire, *Geol. Mag.*, vol. 74, no. 882, pp. 534–553, 1937.
12. Thompson, Warren O., Original Structures of Beaches, Bars, and Dunes, *Bull. Geol. Soc. Amer.*, vol. 48, pp. 723–752, 1937.



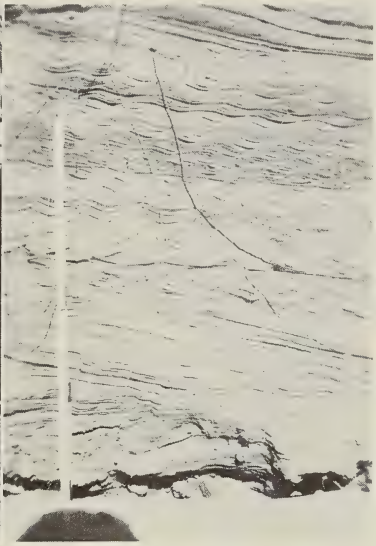
Æolian cross-lamination, Coconino sandstone.



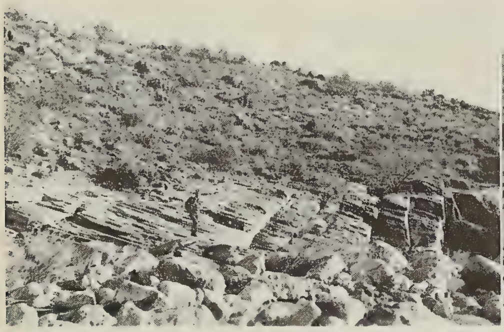
Modern dune cross-lamination, Lake Mead.



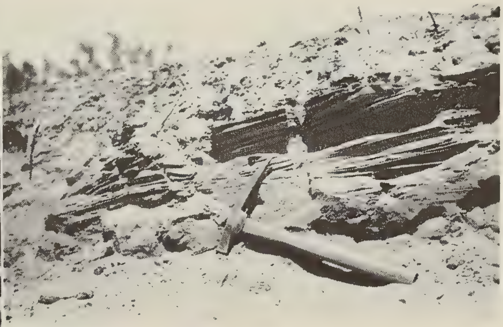
Devonian ripple cross-lamination.



Colorado river delta ripple cross-lamination.



Delta-front lamination, Supai formation.



Delta-front lamination, middle Colorado River delta.

EARTH HISTORY OF A PORTION OF THE PACIFIC NORTHWEST

JOHN P. BUWALDA
California Institute of Technology
Research Associate, The Carnegie Institution of Washington

INTRODUCTION

Of some parts of the land areas of the earth the geological history is readily decipherable, and a long series of physical events can be demonstrated, typified by such changes as submergence below or emergence above the sea, vulcanism, uplift and resulting deep dissection and eventual complete erosional planation, mountain building with development of very complex structure in the crust, and broad down-bowing or down-faulting of districts to form basins in which bodies of sediments often thousands of feet thick accumulate. Such episodes constitute the physical history of the earth. But in large areas the bedrock of the country, from which the history must mainly be read, is covered by alluvium, or lava flows have spread over it to the extent of thousands of square miles, or glacial detritus has masked the underlying rocks and structures. The geological history of such concealed parts of the crust was until quite recently almost entirely unrecoverable, and though great advances have been made in drilling and geophysical methods during the past decade or two, the accessible record of physical events in such areas is even now a partial rather than a complete one. Examples of districts in which the historical records can be made out with facility are the Coast Ranges of California and the Grand Canyon of the Colorado River; in striking contrast are the deeply alluviated floor of the San Joaquin Valley, the wide prairie plains of the interior of the continent, and parts of the northeast quarter of North America deeply encumbered with glacial waste.

Another such region is the subject of this essay. A vast area often spoken of as the Columbia lava plateaus, and comprising roughly the southeastern quarter of Washington, much of the eastern two-thirds of Oregon, and parts of southern Idaho, experienced in Miocene time tremendous outpourings of dark-colored basaltic lava, and over most of this area this formation now forms the country rock, produces the soil through its decay, and gives the landscape its dark brown and other somber tones. The successive flows of lava covered the country rather completely, except for occasional hills or mountain ridges which were not buried by them, and consequently concealed effectively the underlying formations from which alone the pre-lava geological history of the region could have been read. In one area however, known as the John Day country, in north central Oregon, post-lava mountain making has buckled and uplifted the crust, has caused streams to cut away great thicknesses of the lava and of the underlying strata as well, and has thereby exposed an unusually complete cross section of the formation down to the oldest rocks. Erosion has thus cut a window through the lava mask, rendering accessible the field data from which by inference and inductive reasoning it has been possible to construct an unusually complete geological history for the province even though most of its area offers few or no historical clues. The geological conditions and history revealed in the window are not to be regarded as local but doubtless are representative in a general way of most of the huge region covered by the lavas, and hence take on added significance as a particular broadly exemplifying the whole.

Since geological history includes such diverse phenomena as mountain making, vulcanism, development and extinction of different races of land animals, marine organisms, and land plants, and changes of climate, the evidences indicating the successive events are drawn from very different kinds of geological and palæontological materials in the field and in the laboratory. And consequently investigators with quite different interests and employing various field and laboratory techniques all contribute to the final narrative of geological events. Such studies are naturally and necessarily carried forward as a cooperative effort, and the elucidation of the geological history of the John Day region is a very striking example of the joining of scientific personnel under the leadership of an eminent authority to prosecute vigorously and effectively the various researches relating to the same general problem. Dr. Merriam initiated the investigation of the geological history of the region just before the opening of the present century, and shortly thereafter published important results relating both to the physical history and to the development of mammalian life. Although nearly four decades have elapsed and much further work has been done in the area, the conclusions expressed at that time have stood unchallenged, and Dr. Merriam's original and pioneer paper has served as the principal guide to the geology of the region down to the present time. Soon after 1920 Dr. Merriam interested some of his former students at the University of California in different phases of the geology and palæontology. Dr. Chester Stock and assistants undertook extensive collecting of vertebrate fossils in the formations of the John Day drainage basin, and in the study of these materials in the laboratory. Dr. Earl L. Packard investigated first the invertebrates of the Cretaceous strata of the Mitchell district, and later with student assistants discovered and studied important invertebrate faunas in earlier Mesozoic and several Paleozoic formations, the presence of which in this part of the West had scarcely been suspected. The writer and student assistants mapped geologically the Mitchell Quadrangle and the Picture Gorge Special Sheet area, which together afford a very productive research project, involving highly cooperative relations between the participating workers, and leading to a great increase and broadening of our knowledge of the geological history of this extensive region.

GEOLOGIC SETTING OF THE JOHN DAY REGION

The western part of the United States is largely mountainous. The rough country lying between the Mississippi Valley and the Pacific Ocean differs greatly in ruggedness however, and although known as the North American Cordillera or continental backbone it is constituted of several definite north-south topographic units. The easternmost of these, the Rocky Mountains themselves, from Idaho and Montana to New Mexico and Arizona, consists both of large areas of very rugged mountains and of high-lying plateaus often surmounted by individual ranges of impressive height and grandeur. One hundred and fifty miles to the west lies the Cascade Range, separating the western third of Washington and Oregon from the eastern two-thirds of those states, and farther south, as a continuation of the Cascade uplift, stands the Sierra Nevada of

California, somewhat more distant from the Rockies. A linear but somewhat discontinuous series of valleys forms a persistent belt along the west side of the Cascades and Sierra, and west of it, along the Pacific Coast, rise the Coast Ranges, from northwestern Washington to southern California. The wide province between the Cascades and Sierras on the west and the Rockies on the east is somewhat depressed relatively to these ranges but lies mainly at elevations between 2000 and 5000 feet. The southern part of this intermountain country is usually termed the Great Basin province, because of its interior drainage, but from roughly the Nevada-Oregon boundary northward it is usually designated as the Columbia lava plateau. The Rockies to the east are built mainly of pre-Tertiary rocks, as are the Okanogan Highlands to the north, and the lavas forming much of the surface of the Columbia lava plateaus apparently flowed against the flanks of these adjoining higher areas. On the west however the lavas pass beneath the younger strata and lava flows of the Cascade Range. In the midst of the Columbia lava plateaus rise the Blue Mountains of eastern Oregon, probably existent at the time of the extrusion of the lavas, but certainly uplifted greatly in post-lava time. Lower mountainous country extends westward from the Blue Mountains; the westward-flowing branches of the John Day River drain this territory and have deeply dissected it. Consequently it consists mainly of east-west ridges, sometimes sharp-crested and narrow but commonly several miles wide, several hundred to more than two thousand feet in height above the streams, and either flattish-topped or dissected and rugged throughout. The valleys separating the ridges usually have sloping sides and either very narrow or no flat floors, the only notable exception being the valley of the upper John Day River, along which the alluvial flood plains and terraces sometimes attain a width of about one mile. Commonly the upper parts of the ridges are lava-covered, but in the lower parts of the valleys and in some of the tributary drainages the older rocks are well exposed, and it is here that the earlier chapters of the geological history can be read.

If the Columbia lavas were cleanly stripped off the entire region it is not probable that the geological conditions would be found to be similar throughout the entire uncovered plateau, but the similarity would presumably be as great as exists within the Cascades and in the Coast Ranges to the west. If this be true, the history determined from the exposures in the restricted area of pre-lava rocks laid bare by the cutting of the John Day and its tributaries can legitimately be regarded as either applicable to the much larger area of the entire Columbia lava plateaus or as at least indicating roughly the nature of the physical events which took place in that province during corresponding epochs of geological time. Such similarity in history is suggested in the limited number of exposures of the Tertiary rocks lying beneath the lavas at the northern margin of the lava field in central Washington, in the core of the Cascade Range as shown in the walls of the Columbia River gorge, and along the northern and southern margins of the western Snake River Valley in southern Idaho.

Earliest Events in the Geological History

Pre-Cretaceous rocks are exposed over only very small areas on the quadrangles mapped in detail by the writer but they have been extensively laid bare to the southeast, on the south slopes of the Ochoco Mountains. The researches of Dr. Earl L. Packard and Dr. Ralph L. Lupher have revealed that during the Paleozoic era the sea apparently swept across central Oregon at least several times, depositing during each incursion hundreds to thousands of feet of limestone, sandstone, and shale, or more often alternations of these. Considering that the strata are as diverse in age as Devonian, Mississippian, and Permian, it is quite certain that following each invasion and deposition uplift occurred and part of the beds laid down were eroded away. The Paleozoic formations discovered and studied by Dr. Packard contain a great variety of invertebrate fossils. They indicate that at least most of the Paleozoic formations were deposited in the sea rather than in fresh-water lakes or stream valleys or desert valleys. They date the formations quite accurately as to time of deposition. The record of some of the principal physical events of Paleozoic time is therefore one dominated by marine deposition during recurrent invasions of the sea. Mountain making presumably closed the Paleozoic, for Mesozoic sediments lie unconformably across the worn edges of the Paleozoic strata.

During the next period, the Triassic, the sea invaded this region at least twice, depositing quite different types of sediments. Conglomerates, limestones, and cherts accumulated during the first incursion and their rocky outcrops in the region now yield marine fossils in such quantity and of such good preservation as to demonstrate clearly both the age of the strata and the marine conditions under which they were deposited. Later in the Triassic, presumably after the earlier sea had been driven out partially or completely by uplift of the sea bottom and part of the limestones and cherts had been cut away by erosion, subsidence again occurred, and a thick body of sediments, mainly muds, was laid down, to be transformed later into dark shales, and still later, as a consequence of mountain-making pressures, into slates. These rocks are well exposed in outcrops west of Mitchell and south of the town of John Day.

The sequence of events during the Jurassic period likewise consisted of at least two invasions of the region by the sea, as inferred from data gathered during the researches of Dr. Packard and Dr. Lupher. During each of these floodings large thicknesses of sandstone, shale, and other types of sediments were brought in by streams from the adjoining Jurassic land masses, to form the rocks now well exposed on the south slopes of the Ochoco Mountains. Many species of ammonites, a group of spiral-shelled creatures which grew to large size and enormous diversity of internal structure, left their remains in these strata, testifying both to the huge expansion and evolutionary development of this family and to the climatic and general environmental conditions in the sea during their lifetimes. Toward the end of the Jurassic the whole eastern Oregon region apparently experienced the same strong mountain-making disturbances suffered by most of the Pacific Coast region. The strata were folded and faulted presumably as a result of compressive forces operating more or less horizontally in the crust. The axes of the anticlines trend northeast and southwest, in contrast with the east-west axes of the folds in the Tertiary formations of the region. It is clear from the divergence of axes that mechanically the mountain building of the later periods was not merely a simple continuation of the Jurassic deformation. Igneous activity also characterized some parts of the Triassic and Jurassic. Basic magmas, now changed to serpentine, were apparently intruded during the Triassic, and more acidic rocks, of the nature of diorite and granite, invaded the crust in this region during the Jurassic.

After all these disturbances a rather long period of erosion ensued, removing vast quantities of the upfolded and probably upfaulted sections of the Jurassic and all earlier rocks. This is known from the fact that the overlying Cretaceous and younger strata all bear strongly unconformable relations to the Jurassic and older formations. Such unconformities are very significant structural features in the unraveling of geological history. Such an unconformity is an old erosion surface, on which the beds above the unconformity were deposited. It therefore records several events. If the beds below it contain marine fossils, a first episode indicated at the unconformity is the subsidence of the region below the sea, in this case in Jurassic time. Deposition of the strata occurred during and following the subsidence, entombing the marine organisms of the time, and from these the geologist and palæontologist determine both the geologic date of the deposition, and also much of the conditions of water temperature, currents, depth of water, and other factors. The fact that the Jurassic strata are tilted more steeply than the overlying Cretaceous beds demonstrates that mountain making occurred between the dates of deposition of the two sets of beds, and the degree of divergence in dip gives some measure of the acuteness or severity of the deformation. In this instance it was severe, for the pre-Cretaceous formations stand at much higher angles than do the Cretaceous and later beds. And of course the nature and fossil content of the beds immediately above the unconformity demonstrate the date and the physical conditions under which these overlying rocks were deposited. The Jurassic–upper Cretaceous unconformity therefore epitomizes several geological events the total duration of which may well have been several millions of years.

The Upper Cretaceous Sea

During the Cretaceous period a great inland sea covered much of the interior of North America, connecting the Gulf of Mexico with the Arctic. Large parts of the Pacific Coast were submerged, but the inundating waters were not a part of the great central seaway, but marginal waters of the Pacific spreading over the western parts of the continent. This sea apparently did not transgress the site of the present Sierra Nevada but it did cover at least considerable parts of eastern Oregon and presumably crossed some part of the area now occupied by the Cascade Range. Upper Cretaceous rocks are exposed at a number of localities in the John Day region, both in the two quadrangles mapped by the writer and in the John Day Valley and the Ochoco Mountains in districts studied by Dr. Packard, who has collected extensively from these strata and determined their age accurately from the palæontological materials. The sedimentary section, several thousand feet thick in the Mitchell area, consists of an alternation of several hundreds of feet of dark shale and intercalated thin beds of limestone and sandstone, and of beds of conglomerate usually not more than a few tens of feet thick. This succession of dark shale and conglomerate is repeated many times. The beds are tilted, commonly to angles of about 30°, and as they are now being eroded down, the conglomerates, being much more resistant, form persistent reef-like ridges continuous for miles across the country, while the intervening shale forms valleys between the ridges and naturally gives smooth grassland topography in contrast with the rugged outcrops marking the crests of the conglomerate ridges. An important fault of quite large displacement separates the Mitchell area of Cretaceous beds into two parts, in which the structure is very different. The date of this faulting is set rather closely by the fact that it cuts the Cretaceous beds but not the overlying Tertiary formations. It presumably occurred therefore at the same time that the Cretaceous beds were folded and tiled. Its trend is not parallel, however, to the general trend of the Cretaceous strata; the fault bears somewhat north of east while the formation strikes northeast or even more northerly. The lines of folding hence tend to suggest a continuation of the deformation evident in the Paleozoic and Mesozoic formations in the Ochoco Mountains to the southeast, while the fault approximates more closely the later axial trends in the Tertiary rocks. The contrast in the directions of folding in the Cretaceous and post-Cretaceous formations is very striking, either as viewed in the field in the Mitchell area by one who has gained acquaintance with the distribution of the formations, or as observed on a colored geological map on which the areal pattern of distribution has been portrayed.

While much of the upper Cretaceous deposits consists of fine sediments, the abundance of river gravels certainly suggests strongly that the shore of the Cretaceous sea was not far distant. And the coarseness of the gravels further suggests that the adjoining lands were sufficiently rugged to produce during erosion great quantities of boulders.

The evidence afforded by study of the John Day region is therefore that part of the intermountain country, perhaps a considerable fraction of the Columbia lava plateaus, was submerged beneath the sea in upper Cretaceous time, that the shore line of the Pacific or of an arm of it lay between the John Day country and the west base of the Rocky Mountains in Idaho, and that the region yielding the sediments was at least somewhat rugged. This history contrasts very markedly with what is to follow in Tertiary time.

Earlier Cenozoic Episodes

Eventful as was pre-Cenozoic time in central Oregon, as outlined in preceding sections, the geological history during the Tertiary and the Quaternary divisions of the Cenozoic was astoundingly varied and complex. A long series of physical changes occurred rapidly and was completed in a period short in a geological sense.

Subsequently to their deposition the upper Cretaceous beds were moderately to strongly folded and faulted. This deformation and the erosional truncation of the strata occurred in late Cretaceous and early Tertiary time, and previous to the deposition of the first Tertiary formation a landscape had been developed which, truncating the Cretaceous beds, was rather smooth or at most somewhat rolling. Probably in the later half of the Eocene volcanic activity began. Lavas were poured out over the country and even greater quantities of explosive products—volcanic tuffs in the form of ash, pumice, and angular lava fragments forming beds of agglomerate. The accumulation grew thousands of feet thick. The activity must have continued for a long period, for some of the lava flows in the series developed decayed upper surfaces before being buried by subsequent deposits. Most of the volume of material is fragmental, and its composition varies from basaltic through andesitic to more acidic rocks, but the bulk of it is of andesitic or intermediate composition. This great series of volcanics was named the Clarno formation by Dr. Merriam. It has yielded no vertebrate fossils so far as known. Dr. Chaney has collected and studied excellent assemblages of fossil plants from this formation. It is regarded as Eocene in age. The climate at time of deposition was much more humid than the present one. The history indicated by the Clarno is therefore one of intermittent vulcanism, much of it explosive in character, with periods of erosion of the varied materials accumulated, until the level of the country had been built up some thousands of feet. The successive landscapes must have been those of a typical volcanic field, except that in all probability a forest grew on the ashy surface a large part of the time.

The John Day formation rests upon the Clarno. It is renowned for the large quantities of vertebrate fossils which it has contributed to museums in this country and abroad. In appearance it is a very striking formation in the field. The Clarno beneath it is often highly colored, as volcanic accumulations commonly are, but the colors are usually dark—reds, browns, greens, and blues. The John Day is commonly light-colored, in shades of green and blue, often white. It sculptures into striking badlands, with high fluted cliffs, columns, minarets, and horizontal shelves on vertical faces. While it contains a large quantity of ash, it is dominantly a sedimentary formation, showing very distinct stratification, and its origin was clearly not dominantly volcanic but by accumulation in a basin. Most of the formation is coarse sandstone, and while water bodies doubtless existed during its deposition, the bulk of the material was probably laid down by streams. The formation is known at many places in eastern Oregon, suggesting that either a large basin of deposition existed in the region at that time (perhaps comparable to the San Joaquin Valley of California except that the source area of the sediments was largely volcanic country), or numerous smaller basins were scattered about but comprised a considerable fraction of the total area.

The vertebrate fossil material derived from the John Day formation indicates that it is of middle to upper Oligocene age and that the upper part may be as young as lower Miocene. It is some thousands of feet thick but its total thickness is nowhere shown so far as known.

The relation of the John Day formation is important in its bearing on the tectonic history. Usually the Clarno volcanics grade up into the John Day sediments, through diminution of volcanic material and increase of ordinary nonvolcanic sediments. Commonly the beds above and below the contact of the two formations are parallel, and no marked erosional break exists; at least the contact is no more striking than many others within the Clarno. Although the relations are therefore usually suggestive of continuous accumulation with mere change in nature of materials laid down, and therefore of close relationship in time, Dr. Ralph W. Chaney has described a locality on the south side of the Ochoco Mountains which appears to him to demonstrate unconformable relations between the two sets of beds. If this be correct, a period of erosion intervened between the accumulation of the two formations and they may be of quite distinct age. The implications of the John Day historically are in any case that the Clarno vulcanism in the region had largely ceased and a long period during the Oligocene was involved in the filling of basins with sediments, in which were entombed the remains of the great variety of land mammals which roamed over the wide floodplains of the streams. Climatically the country was still humid, though perhaps less so than during Clarno time.

The John Day and the underlying Clarno are moderately folded, with dips of 20° to 30° very common. The degree of deformation of the two is not very different, suggesting that little mountain making had occurred after Clarno and before John Day time. These formations were folded along east-west axes, in contrast with the northeast trends of the underlying Cretaceous beds, and the implication is that the direction of the application of force causing the folding changed in early Tertiary time. Erosion beveled the rising folds of John Day formation and cut away a large part of the total volume of the beds before the next younger formation—the Columbia lavas—came to rest upon them.

LATER CENOZOIC HISTORY

By middle Miocene time erosion had developed a rather subdued landscape over most of the John Day country, if one may judge from the nature of the contact of the Columbia lavas with the underlying formations. This landscape had been cut largely across Clarno volcanics and John Day sediments, but probably patches of Cretaceous and even older formations were then already exposed in the region. Out on this surface poured successive flows of very liquid basaltic lava. They issued from fissures which may still be seen, filled with the black rock resulting from the congealing of the liquid lava which filled the long vents when outflow ceased. They are of course visible now only where the flows which they fed have been stripped off the older rocks through which the fissures passed, or in the faces of bluffs made of Columbia basalt flows, in which fissures passing through the lower horizontal sheets of lava can be seen filled with lava which at the time of freezing was on its way upward to feed the then surface flow. The process was clearly a quiet one, as explosive products are virtually absent from the lava series. The flows, most of them many miles long, of very uniform thickness throughout their visible length, and ranging from as little as 10 feet to over 200 feet in thickness, accumulated to a maximum depth of about one-half mile in some parts of the John Day country. Trees often grew on the surface of one flow before the succeeding one pushed them down, encased them in the lava, and sometimes permitted them to be completely petrified, to be preserved to the present day. The Columbia lavas record a stage in the history characterized by long-continued quiet upwelling of vast quantities of very liquid basalt, flooding the whole country, blotting out most of its pre-lava relief, and giving to a region of some diversity in landscape and surface geology a monotony in topography and general appearance during middle Miocene time possessed only by the upper surfaces of dark-colored lavas.

Before the last basaltic flow of the series had been extruded, light-colored sands, shales, and ash beds began to accumulate on the surface and hence the upper basalt flows are locally interbedded with a sedimentary formation of middle Miocene age named by Dr. Merriam the Mascall. These strata were deposited horizontally on the lavas while the latter were still level and undeformed, so that the two are still parallel. No unconformity separates the formations, and hence the age of the Mascall defines the age also of the Columbia lavas. They are middle Miocene, as is indicated by the wealth of vertebrate fossil material which the Mascall has yielded. The Mascall was apparently not deposited over very large areas of the Columbia lavas; perhaps it was laid down in a gently warped basin in the surface of the lavas.

Some time after the deposition of the Mascall further folding along east-west axes commenced. Along John Day Valley the formation has been preserved in a syncline. The river had planed off the upper part of the Mascall section and had opened a broad valley, the floor of which was mainly Mascall and the sloping sides of which, rising some hundreds of feet above the floor, were mainly Columbia lavas dipping beneath the valley from both sides. The river left the synclinal valley at Picture Gorge and crossed other fold axes. In middle Pliocene time further uplift apparently occurred athwart the course of the stream, causing it to lose velocity in John Day Valley and to deposit several hundred feet of gravel and ash beds. This is the Rattlesnake formation, also containing abundant vertebrate fossil remains. Its fauna, studied by Dr. Merriam and Dr. Stock, indicates a middle Pliocene age, dates some of the later folding, and, like the Mascall fauna, demonstrates that the climate of the region was more arid than the climates of the early part of the Tertiary. During the accumulation of the Rattlesnake beds a rhyolite flow coursed down the John Day Valley, and is interstratified with the formation, like the rhyolite flows in the John Day sediments. The igneous history, considering the variety of types of lavas which were extruded at different dates, was indeed remarkable. Subsequent to the deposition of the Rattlesnake further down-bowing of the syncline occurred, sufficient to give the beds a gentle dip from both sides of the valley toward the middle. Then John Day River proceeded to cut its valley deeper and well down through the Rattlesnake, which it exposed widely along the valley sides. In its downward progress it has alternately cut and filled, and hence stream terraces have been well developed along the sides of the valley and are exhibited excellently even at the mouth of Picture Gorge, where they have been cut in the hard lava. The present chapter in the history of the John Day country is one of rather vigorous dissection and erosional degradation.

Origin of the Present Landscape

The present landscape of the John Day region owes its character to several different factors. The height of the region above the surrounding lava plateaus results from uplift in relatively recent geological time, and this process has presumably not been completed but is continuing. It has occurred repeatedly during the past several tens of millions of years. The details of the topography, the ridges, valleys, and canyons, were formed by stream dissection and tributary erosion. But here the structure influenced the result greatly. The main branches of John Day River flow westward parallel to the axes of folding, and while they are forced to cut across these axes from time to time, the stream pattern was largely determined by the folds. Large flat uplands are merely the stripped upper surfaces, in most cases, of the Columbia lavas, which are resistant to erosion. The Rattlesnake, Mascall, and John Day formations, and to lesser degree the Clarno, tend to give badland topography, owing to their lack of resistance to the erosional agencies.

At a few localities in the John Day region patches of old surface exist which appear to be referable to a prolonged stand of the region at one level, during which it was worn down to a very subdued topography. Uplift raised these areas to a level at which they will be dissected eventually, but they still persist as evidence of a period of smooth planation.

Man's Environment Largely Geologically Determined

The natural elements in an environment which largely control man's activities are the topography, soil, climate, water supply, vegetation, and mineral wealth. It is interesting to observe to what degree these factors are geologically determined in the John Day region, the subject of this essay.

The topography is too rugged for extensive agriculture; it lends itself to grazing only. As has been stated, the form of the landscape is the result of geologically recent uplift and the dissection which streams habitually undertake in areas raised above the surrounding country.

The soil, derived from the underlying rocks for the most part, partakes of their nature in a hilly country, and is hence geologically determined. The folding and the dissection have governed the areal distribution of the different types of rock and hence to large degree also the different kinds of soil. The valley soils were of course accumulated by stream action, which is geological.

The climate is in part a function of latitude and of distance from the ocean to the west. But the altitude of the country is important, and uplift, a geological process, determined that. And the great barrier of the Cascade Range to the west, shutting off the moisture which normally would be available from the Pacific, and its equable climate, owes its height to mountain building, a geological activity.

Water supply depends upon climate locally, and upon topography and underground water conditions. These are all related to geological factors. And vegetation, depending upon climate, soil, and water supply, is indirectly governed by geological considerations. Mineral deposits, sparse here, are of course strictly geological products whose occurrence is determined by the past history of the region. It thus appears that the use to which man can put this region is influenced very greatly by the geological history which it and closely adjacent areas have experienced.

Certain Inferences

The eventful geological history of the John Day region has been set forth at some length but by no means in the available detail. It suggests certain generalizations.

The rock materials of which the region is built, and the structure or plan according to which they are put together, are the result not of an episode or a process acting suddenly, but of a long series of events of quite diverse types. Many geological events contributed to the formation of the crust as we now see it in the eastern Oregon country. The movement has been an evolutionary one in the sense that a long series of changes produced a certain result.

The present landscape is only one in a long series of landscapes which have existed in this region, many of them very unlike the contemporary one. Great changes in relief, in stream pattern, in climate, in vegetation, in ruggedness have altered the view from period to period. And with the processes continuing it is clear that additional landscapes of different types will be produced in the future, and that the present one merely links a past with a future series.

From the nature of the formations, the structures, unconformities, and the fossils of former plant and animal life it is clear that the processes operating in past ages in the John Day country were not greatly different from those now or recently active in the area. Uplift, depression, stream work, sedimentation, erosion, vulcanism, the development of new types of animals and plants and the extinction of earlier species have recurred continuously. They are operating now in or near the area.

The present is but a moment in geological time, between an almost unbelievably long past and an indefinitely long future. In this long physical and life history the present is not a special moment so far as that history is concerned, though there is a tendency in our thinking to make it so.

But from a human standpoint it is interesting that, considering the hundreds of millions of years that the geological features of this region have been in the making, it has been only during the past hundred years that man has been able to comprehend them.

PHILOSOPHY AND INTERPRETATION

COOPERATIVE RESEARCH, ITS CONDUCT AND INTERPRETATION

FRANK F. BUNKER

Editor, Office of Publications
The Carnegie Institution of Washington

IN ORDER that the results of the systematic and directed effort to obtain new knowledge which is called research may be made generally understandable and widely available restatement and reiterated explanations are necessary. A survey of the practice and procedure followed in research by the Carnegie Institution will disclose the reasons why investigation must be accompanied or succeeded by exposition and interpretation if newly developed concepts are to become quickly embedded in the common knowledge of mankind.

I

There are in the United States today many hundreds of agencies engaged wholly or partly in research. They comprise laboratories, developed and maintained by industrial establishments for study of highly specialized problems which originate in manufacture of their respective products; museums which were established primarily for educational purposes but which in a number of instances have added lines of investigation related to or suggested by their collections; universities, many of which look upon the research function as being almost as important as the teaching function; various state and federal departments and bureaus, each occupied with problems of practical concern; and, lastly, organizations founded expressly for conducting research either in definitely restricted fields, as with certain medical associations and societies, or in areas of operation of unlimited range.

To the last of these groups belongs the Carnegie Institution of Washington, whose staff, now consisting of 150 full-time investigators and 60 research associates, has devoted itself and all the resources at command of the organization to the prosecution of research under injunction of the Founder: "to expand known forces, to discover and utilize unknown forces for the benefit of man." For thirty-seven years the Institution has been at work upon fundamental problems in many fields but chiefly in the physical and biological sciences although mathematics, the social sciences, and the humanities have not been neglected.

Founded two years after the turn of the present century, the Institution began its work at a time when scientific research stood at the threshold of a period of unprecedented fruitfulness. In the magnitude and brilliancy of man's achievements during age-long effort to learn more about the world—about its essential constituents, about the way it functions, about the laws that govern it—the elapsed period of the twentieth century surpasses all other periods of like duration.

Although basic discoveries in many branches of science were made by early workers, the scientists of the twentieth century have exhibited extraordinary resourcefulness in developing new and more powerful tools and techniques, in consequence of which the pace of discovery has been greatly accelerated. Moreover, the scientists of the period have been extraordinarily successful in evolving synthesizing generalizations that make it possible for the isolated facts of discovery to be seen in their true and significant relationships. Fresh discovery and fresh generalization, in turn, have brought changes in man's concepts so sweeping that repeatedly he has been compelled to fashion anew his thought-models of the universe.

Preeminent among the achievements of this century stands the group of brilliant discoveries leading to the view advanced by Albert Einstein very recently that what impresses our senses as matter is really a great concentration of energy and that grounds exist for expecting the physics of the future to demonstrate that the universe is built up solely of fields of energy.

The investigators who have done such remarkable work on the structure of the atom are concerned with dimensions that are incredibly small, of the order, indeed, of the billionth part of the dimensions of the smallest object which can be measured with the highest-powered microscope yet constructed. Equally astonishing have been the achievements of the astronomers who stand at the opposite pole of physical magnitudes and deal with dimensions that are superlatively large.

Present-century research in this field, for example, has pushed the limits of observable space out to a distance so great that it would take light, which travels at a speed of six thousand billion miles a year, five hundred million years to traverse it. It has also shown that within the space thus defined there are probably a hundred million galaxies similar in essential respects to our own galaxy of the Milky Way. Moreover, with present tools and methods, the distances, diameters, and general properties of these galaxies can be determined with considerable definiteness out to ten million light-years.

The eventful progress in physical science made during the present century has its counterpart among the sciences which have to do with the visible and invisible world of living things, leading, not only to the overturning of concepts once thought to have been enduringly established, but also to the development of entirely new divisions of biological science. And everywhere, in every branch of science, whether of those dealing with inert matter or with life, research is revealing unsuspected connections among phenomena that look to be unrelated, unsuspected simplicities within apparent complexities, and unsuspected orderliness amid seemingly bewildering diversity. Indeed, it is, perhaps, this very orderliness of the universe that constitutes the supreme discovery of science, an orderliness that cannot fail to move profoundly him who contemplates the wonder and glory and beauty of it all.

Two factors go far toward accounting for the great outburst of research activity characteristic of the twentieth century:

First, following Mr. Carnegie's lead, rich men and women and rich business organizations, as never before, have made it possible for persons with aptitude for research to devote their entire time to investigation and to conduct their work with instruments of the greatest precision and power and with facilities that are ample.

Again, as never before, there is team work among investigators who recognize neither national nor group boundaries. Neither are there "trade secrets" in scientific research. Progress reports are continually being interchanged among the workers throughout the world and thus each shares what he has learned with all the others; so, also, each borrows from others and builds upon what he borrows.

An army of investigators, then, trained for the tasks, relieved of economic worry, supplied with every facility required, and working in tacit cooperation, accounts for the epoch-creating achievements now being made in the various fields of scientific research.

It was at the beginning of this quickening period that the Carnegie Institution of Washington was established. As a single research agency, though itself of composite organization, it has taken place in the ranks alongside the other institutions, societies, and agencies that are endeavoring to fathom the mysteries of the universe. In this effort the Institution, through its departments and divisions, has shared in every major advance that has been made in the physical and biological sciences during the present century. Indeed, it is no overstatement to say that in many branches of research the Institution has exercised definite leadership in discovery and integration.

Because of the experience in research which the staff of Carnegie Institution has had, experience acquired under wise and farsighted guidance, because of the fundamental nature of the problems for which solution by it has been sought, and because of the fact that the Institution has achievements to its credit in virtually every field of investigation, the mode of procedure which has developed among its investigators may be examined with profit. In the attempt to learn what the essential features of the research procedure are, as now pursued by Institution workers, it is fitting to examine the line of action which has been taken in the prosecution of specific projects.

II

As a representative example, it will be instructive to review the course followed in that group of researches having to do with photosynthesis, as conducted by Dr. H. A. Spoehr, the present chairman of the Division of Plant Biology, and his associates.

As now viewed, the cells in the green leaves of plants are chemical laboratories of microscopical dimensions. To these laboratories the mechanism of the plants brings air which normally contains about 0.03 per cent of carbon dioxide and weak watery solutions from the soil. Within the cells are minute bodies, the chloroplasts, which contain green and yellow pigments. These chlorophyll-bearing bodies, when activated by sunlight, have the power of converting carbon dioxide and water into glucose. In the process 6 molecules of carbon dioxide combine chemically with 6 molecules of water and produce 1 molecule of glucose and a by-product of 6 molecules of oxygen.

Derivatives of the glucose—the sugars, starch, fats, and protein substances, which the plant also has the power of producing—are stored in its leaves, stalk, roots, fruit, and other tissues against the time when reserves are needed. These substances represent caches of chemical energy, utilizable by animals and man, created through conversion of the radiant energy of the sun. It is these caches of energy that constitute not only the basic food supply of all the life of the world, including that of man, but also the great deposits of coal, occurring in all gradational stages of carbonization, and also perhaps oil and gas, that are so essential to man's comfort and efficiency. The molecules of oxygen, released when glucose is formed, pass into the air thereby replenishing the general store of oxygen which is constantly being depleted by respiration and the other forms of combustion.

This process by which green-leaved plants, with the aid of sunlight, convert inorganic compounds into organic carbohydrates, giving off oxygen the while, is called photosynthesis. As the name suggests it represents a synthesis or putting together with the aid of light.

There is a reverse process in which the energy required by the plant to keep alive and to grow is obtained through the combustion of its own carbohydrates. In this process, the process of respiration, as with animals, oxygen is consumed and carbon dioxide is released. Doubtless respiration is taking place all the time, but in the sunlight it is so completely overshadowed by the much more active counter process of photosynthesis as not to be noticeable.

Study of the subject by Carnegie Institution began with announcement in the Year Book of 1910: "Dr. H. A. Spoehr, of the University of Chicago, has been appointed a member of the staff. . . . He will devote his attention to a group of problems in chemical physiology." In the Year Book of the following year (1911) is to be found Dr. Spoehr's first progress report, entitled "Chemical Effects of Radiant Energy in Plant Processes." Annually since then a published report on the progress of the study has appeared.

As with the inception of all the major research projects undertaken by the Institution, commencement of the study of photosynthesis was not a leap in the dark. There was a foundation structure that had been built by others. Thus, during the quarter century ending with 1804 experimental scientists beginning with Joseph Priestly in England, who first isolated oxygen, and including workers in Sweden, Holland, France, and Switzerland, had, through successive stages, arrived at a correct view of the function which green plants subserve in the economy of nature and had recognized also the major features in the phenomenon of photosynthesis and the associated process of plant respiration.

The next advance in study of photosynthesis related to attempts to learn how the plant actually operates in the manufacture of its carbohydrate products and their derivatives. However, such is the difficulty of tracing the steps taken by the living plant in this process, and of determining the exact chemical reactions that occur within the plant, that little could be accomplished in this direction until chemistry as a science had developed a more assured technique with a greater body of verifiable knowledge, especially knowledge relating to the carbon compounds.

Following demonstration by Friedrich Wöhler, in 1828, that one substance, urea, hitherto found only in living matter, could be made in the laboratory, attention turned to the synthetic preparation of natural products. Gradually a number of substances produced by living organisms were isolated and their constituent elements identified. Thus, by the close of the century, the three great classes of organic compounds, the proteins, fats, and carbohydrates, had been recognized and some progress had been made toward determining the nature and even toward the synthesis of many of the complex products of living processes.

With the opening years of the present century interest swung back sharply to the green plant and to nature's processes as represented therein. This interest was given impetus through the classic experiments of Richard Willstätter, a German scientist, who, in 1902, began intensive study of the nature and function of chlorophyll, the substance long recognized as constituting one of the essential features of the photosynthetic mechanism. He broke it down into its molecular parts and showed that it is not a single green pigment, as was thought, but apparently an association of two closely related green pigments, chlorophyll *a* and chlorophyll *b*. Willstätter and Stoll, one of his coworkers, not only made a notable contribution to knowledge of what the essential leaf pigments are, but also devised a method by which these pigments can be extracted from the plant and separated from one another for laboratory study.

This was the point reached in study of photosynthesis when Dr. Spoehr, in 1910, began his investigations; this was the level from which the Institution's researches in the subject took off. Examination of the progress reports as recorded in the Year Books of the Institution gives clear indication of the trail taken by this group of explorers as they slowly, laboriously, but surely, penetrated an unknown region. Some of the considerations that influenced the course pursued, as recorded in these reports, follow.

The work began with extensive experiments in photochemical reactions similar to those supposedly occurring in plants and involving such questions as: the rôle of organic acids in the plant leaf and the relation of these substances to protein synthesis; the effect of ultraviolet radiation on carbon dioxide reduction; and the exposure of plant acids, under various conditions of solution, to sunlight for long periods to see whether the resulting products were such that from them sugars could be formed.

Through rigorous experimental study extending over a period of years it became evident that photosynthesis was such a complex phenomenon that further insight into it could be gained only through intensive study of the process with the external conditions so arranged that such factors as temperature, light, air supply, and water vapor could be regulated at will.

After extensive preliminary experimentation, apparatus suitable for the purpose was devised and, thus equipped, study was begun of such matters as: the basic metabolism of leaves and the influence of this metabolism, and of the varying sugar content of the leaves, which the investigators induced, on the photosynthetic process; the rôle of the proteins, amines, and amino acids in both photosynthesis and respiration, in the light of their regulatory and catalytic properties as components of cell protoplasm; and the behavior of the inorganic nutrients of the plant, the mineral salts, in respiration and photosynthesis.

Work along these lines soon brought realization that no clear explanation of the absorption of solutions by cell masses or of the shrinkage that takes place in them owing to the loss of water by osmosis could be had without considering certain fundamental mechanical features of the cell structure. This realization, in turn, led to a series of carefully planned experiments designed to throw light on the nature and function of the principal cell parts and of the multiple and varying structures of the cell protoplasm.

Again, as the investigations progressed, evidence accumulated to support the view that an internal factor exists which plays an essential rôle in the photosynthetic process and which operates quite independently of temperature, light intensity, pressure of the carbon dioxide, and water content of the leaf. Failure of all attempts to reproduce the process outside the living cell as well as direct experiments with living leaves pointed to such conclusion.

Search for such a factor led to interesting efforts to determine whether photosynthesis is intimately associated with the vital activities of the plant, being thereby dependent upon general protoplasmic metabolism, or whether it can proceed independently of these activities, merely contributing the material essential to oxidation and to growth. It was reasoned that if direct interdependence exists between photosynthesis and the metabolic activities of the plant the factors influencing respiration would also affect the photosynthetic activity.

Thereupon an extensive investigation of the respiratory activity of leaves was undertaken in order to identify the more essential factors which affect the rate of the oxidative breakdown of the carbohydrate fuel of the plant and to determine the relation which these factors bear to photosynthetic activities. By 1922 Dr. Spoehr was able to announce the completion of these experiments. His report for the year includes a detailed description of the method and materials employed and the important deductions to which the study led.

Passing over the investigations of the next five years without comment we come now to 1927, when intensive study of the pigments of the chloroplasts was begun, a study which proved to be one of the most difficult of all those relating to the photosynthetic mechanism and method and yet one in which the investigators perhaps achieved their most striking success.

Nearly twenty years before, as already mentioned, Willstätter and Stoll had demonstrated that chlorophyll consisted apparently of two green pigments, which they called chlorophyll *a* and chlorophyll *b*. They also devised a method for extracting these pigments from the plant. Following announcement of the results of their work these pigments were made the subject of many investigations by plant physiologists and organic chemists; however, such are the difficulties presented by the subject that little was actually known of the fundamental physical and chemical properties of these pigments at the time that Dr. Spoehr and his assistants took up the study.

The first problem to be solved by the Institution group was that of providing for a supply of pigments adequate for laboratory needs. This involved testing a great number of plant species to find the ones which could be drawn upon most profitably, and also developing an equipment for handling large quantities of plant material since the yield of pigments at best is extremely small. Then came laborious and painstaking effort devoted to refining the techniques for extracting the pigments and separating them in order that the properties of each might be studied.

As finally developed the most sensitive of the methods employed is the one which is based upon the affinity which certain adsorptive substances, such as magnesium carbonate, lime, and alumina, have for the respective pigments contained in a given solution. Thus, when a solution containing pigments of a leaf extract is filtered through a vertical column of powdered magnesium, as in a glass tube, each pigment in the mixture lodges at a different zone in the column as indicated by a band of distinctive color. In such a stratified column, called a chromatogram, the green top zone is chlorophyll *b*; the blue-green second one is chlorophyll *a*; the yellow third and fourth zones are xanthophylls, zeaxanthin above and lutein below; while near the bottom of the column are two more yellow zones, beta-carotene, next in order, followed by alpha-carotene. Moreover, by use of solvents, the individual pigments can be removed from the powder and crystallized in pure form.

So sensitive is this method of segregation by adsorption that with the use of columns of very small diameter, about 1 millimeter, it is possible to separate as little as 0.0015 of a milligram of carotene from carrot roots into alpha-carotene and beta-carotene. Separations of equally small portions of the xanthophylls have also been made and the pigments identified by readsorption with known xanthophylls.

When a column of segregated pigments is examined by transmitted light, differentiation between the green and yellow groups is apparent at once. The differences among the several yellow pigments are much smaller and can be determined accurately only by means of a spectrophotometer. These differences in the absorption of light as represented by the various pigments are indicative of differences in the chemical structure of the molecules comprising the respective pigments.

As study progressed it became evident that a clearer understanding of the rôle played by the yellow pigments was necessary, which meant, first of all, that a thorough knowledge of their chemical nature should be obtained. Investigation of the chemical constitution of the carotenes and xanthophylls, therefore, was undertaken and this investigation, begun ten years ago, is still in progress, with gratifying results to its credit, it should be noted.

Additional interest has attached to these yellow pigment compounds because of the fact that from them are derived substances of great physiological importance to man, such as vitamine A and, apparently, some of the hormones. Indeed, study of this group has made possible significant advance in the science of nutrition, resulting in application not only to human nutrition but also to many branches of agriculture, such as dairying, animal husbandry, and poultry raising.

When work on the yellow pigments was begun it was generally thought that this group contained but two, carotene and xanthophyll. As the matter now stands it is known that plant leaves, as well as many other substances, contain two principal carotenes and indication of at least three more; also that they contain in addition at least twelve different xanthophylls. Most of these compounds have been isolated and their physical properties and molecular structures studied. Analyses of the crystalline pigments have demonstrated that the carotenes are unsaturated hydrocarbons and that the xanthophylls are oxygen derivatives of the carotenes. Although composed of only two elements, carbon and hydrogen, with oxygen in the case of the xanthophylls, all the pigments of the two groups are extremely complex so that determination of their chemical nature has been an unusually difficult and time-consuming undertaking.

Even though much progress has been made in study of the physical and chemical structure of both the carotenes and xanthophylls very little exact information has been obtained regarding the behavior or fate of these substances in the metabolism of the green leaves. The methods of the chemist have thus far failed to reveal the mechanism of a single transformation of the carotene and xanthophyll pigments in the plants. Nevertheless, some advance has been made toward the understanding of these complex reactions. Many of the conditions affecting the formation of the pigments have been determined, and the nature and relative proportions of these pigments in plants at various stages of development have also been established. The investigators are now hopefully engaged in the effort to develop effective methods of attack in this perhaps the most difficult phase of the photosynthetic problem.

Although the photosynthetic section of the Institution's Division of Plant Biology has devoted much time during the past ten years to study of the physical and chemical structure of leaf pigments and the function of these substances in life processes, other important phases of the photosynthetic process have not been neglected. One such investigation, having to do with the mechanism whereby the leaf absorbs carbon dioxide from the atmosphere and utilizes it in its further activities, is opening up an interesting field for further work. This study, begun a number of years ago but discontinued, was resumed in 1935 with different methods.

It had been realized for a long time that the absorption of carbon dioxide from the surrounding medium by the chlorophyll-bearing tissue constituted the first step in the photosynthetic process and that there must exist some mechanism by means of which this is accomplished. As the concentration of carbon dioxide in the air is very small, it was thought remarkable that the leaf can absorb and assimilate so rapidly such relatively large amounts of this gas. However, the attempts to account for the phenomenon were not convincing, so Dr. Spoehr and his assistants undertook study of this problem.

From a long series of experiments they reached conclusion that carbon dioxide exists in various degrees of fixation within the leaf, indicating that the latter may contain compounds that act as carbon dioxide reservoirs from which the gas may be released with varying degrees of facility. Continued study of the matter showed that a large proportion of the absorption by the living leaf can be ascribed to its water content. On the other hand, the experiments also showed that in the leaf there is an excess of carbon dioxide over the amount calculated to have been taken up by the water.

In point of fact it has been demonstrated that plant tissues are capable of absorbing carbon dioxide in quantities considerably above that accounted for by the solubility of the gas in the water of the tissues. In a report covering this investigation recently issued by Carnegie Institution the probability is suggested that the calcium and magnesium compounds, which are among the basic constituents of the solid residue of plant leaves, play an important rôle in this the first step of the photosynthetic reaction. Doubtless the further study of this problem which will be given it will clear up this and other little known factors of the photosynthetic phenomenon.

III

The foregoing sketch covers the program and procedure of one group of Institution investigators as the work unfolded during a twenty-eight-year period of continuous effort. While far from exhaustive in respect to details, nevertheless the outline is full enough, it is believed, to provide basis for the making of certain significant observations on the way in which scientific research is conducted in the Institution.

It is to be noted, first of all, that this research on photosynthesis is carried forward as a group activity. The personnel set-up, typical of all the research groups of the Institution, consists of a number of skilled investigators directed by one possessing qualities of stimulating leadership, ability to conduct rigorous experimental investigation of high order, and complete familiarity with what is already known and what has already been achieved in the field or fields which the research touches.

In the earlier years of the Institution individuals received short-term grants to assist them in study of specific projects much more frequently than in recent years. Although assistance is still extended to individuals the trend in Institution practice, particularly during the period of President Merriam's administratorship, has been toward group organization in the prosecution of research.

Cooperation among members of a unit group within departmental boundaries; cooperation among unit groups of different departments within the Institution's organization; cooperation among the departments themselves; and cooperation among Institution groups and groups of other agencies, foreign as well as domestic, represent the administrative policy of the Institution in respect to conduct of research, a policy which has been fostered and greatly developed under the leadership of President Merriam. Though not underestimating the value of the results obtainable by able individuals working alone, President Merriam has felt that the problems of research as now envisaged are so complex and difficult and lead the investigators into such disparate branches of science as generally to require united attack by specialists who in their training and viewpoint can bring to bear on the questions raised the various applicable disciplines which they individually represent.

The best short statements of President Merriam's thought in respect to need of organized effort in research are to be found in the Year Books of the Institution. Thus, writing in 1931, in reviewing the thirty-year period of the life of the Institution that had then elapsed, he refers to the matter, saying:

> The changes in objectives, organization, and plans for use of results of the Institution's work as viewed over the three decades of its existence show increasing clearness of vision as to importance of the work undertaken, and as to definition of objectives for researches in an institution of this character. The organization of the Institution has changed from that of slightly related individuals or groups to an agency with distinct unity. Along with advance in effectiveness of the group as a whole, there has been increased advantage to the individual investigator.

Again, in his 1933 report, he returns to the subject in the following statement:

> In many and unexpected ways new and complex problems of this particular period have shown the need, not only of experts concerned with extremely specialized subjects, but along with these the requirement for investigators who can bring knowledge across the boundaries from one field into another in which it may

have relatively high value. Many problems which focus in economics or government, but arise through consideration of special questions in the natural sciences, require this close cooperation of workers in different fields.

In his report for the following year (1934) President Merriam again discusses the trend of Institution practice away from reliance upon individual effort.

By definition of the situation, a research institution will give evidence of progress through its additions to knowledge. At times these advances may be represented by sharply separated individual items. In the greater number of instances many interrelated factors will be involved, of which neither the individual nor the collective significance can be determined for a considerable period. In the larger groups, or departments, established for attack on major questions, a large part of the result may depend for its value upon relation to other contributions. So it may appear that efforts of individual investigators have their largest value at a particular moment when seen as parts of an advancing stream of knowledge. Under such conditions we might be aware of progress and yet it would be difficult to describe the situation in terms of a report.

With reference to definition of its projects, Carnegie Institution has shown rapid changes in method of organization and in what might seem to be objectives. Mr. Carnegie's desire to find the exceptional man and to give him opportunity was in itself a great stimulus in formulation and organization of research activities. As a natural result, early activities of the Institution rested in relatively large measure upon individual efforts and specific projects of limited scope. But the wider range of studies organized by committees set up on request of the Institution led soon to larger projects, and ultimately to establishment of departments such as the Geophysical Laboratory, Marine Biology, and Mount Wilson Observatory. These groups were entirely independent of one another, and in general not intimately related to other organized activities.

The setting up of departments or groups for research on major projects may have appeared to some as a departure from the original idea involving selection of exceptional individual talent and furnishing special opportunity in the fields thus defined. From another point of view, we see that with advance of science there are many types of problems, in which knowledge can be furthered only by cooperative endeavor of many individuals, sometimes in many localities and often operating over periods amounting to several decades. However narrowly confined the problem for such an investigation may seem to be, attainment of the goal generally involves many shifts of procedure, such modifications being dependent upon personal initiative of individuals in the group engaged on the problems. It is also important to appreciate the fact that, for what may be looked upon as loss of personal initiative on the part of those engaged, there may be large compensation through aid of the whole group in support of investigation on the specific parts of the work in which individuals desire especially to advance knowledge.

Finally, in his Year Book report for 1935 President Merriam once more discusses organization for research. Excerpts from this report follow:

> As has been indicated in many earlier reports to the Trustees, the work of the Institution was initiated with numerous relatively small grants which were commonly made available to individuals. By this means it was possible to select both outstanding projects which needed special support, and exceptional individuals whose ability and interest it was important to have turned toward constructive study. . . .
>
> The Institution of today differs in some measure from the Institution of early years in administrative structure, in method of work, in program of research, and in relationship to educational and investigational agencies as well as to the general public. Yet the changes that have ensued are but progressive responses to altered conditions, on the one hand, and to slow but steady conquest, on the other. . . .
>
> But what seemed to be a fundamental shift in program leading to support of larger projects, and to the cooperation of considerable groups of individuals, has in some respects tended to emphasize significance of the exceptional man more than might have been done by continued use of the smaller grants to individuals conducting their researches independently. Organization of departments, and ultimately of divisions, while appearing to require sharper definition of limits for individual activities, has served to bring about concentrated attack by individuals having specific interests in common problems the solution of which could be obtained only by such cooperative effort. In some respects the results have brought increased opportunity for the individual with exceptional vision and leadership, and at the same time have increased the possibilities of advance in research for others with intense enthusiasm and showing exceptional devotion to research. . . .
>
> In still another way this plan of organization has furnished means for satisfactory handling of those more sharply individualized projects known as Minor Grants, corresponding to the means of aid for research mainly utilized in the early stages of the Institution's activities. This support was given with expectation that it would either assist the exceptional individual, or advance study of an exceptional problem investigated by students of recognized ability. As the number of these grants increased, with passing of the years some researches shifted quite naturally to new aspects of investigation, and the problem of keeping touch with work of this type became increasingly difficult. With a divisional program of administration, large value has been found in the relation of Minor Grant projects to the divisions, thus giving not only the aid of special funds, but with this the support also of a group of workers interested in various aspects of the projects undertaken.

This movement toward tightening the Institution's administrative organization for more effective research led to the coalescence of sections and smaller units of the staff which theretofore were independent and unrelated. Thus in 1928 the workers on the various research projects in the plant sciences which the Institution was supporting were combined forming the Division of Plant Biology. This merger included groups working in photosynthesis, in taxonomy, in ecology, in palæobotany, in tree-ring records, and at the Coastal Laboratory, Carmel, California, and the Desert Laboratory, Tucson, Arizona.

In 1929, the Division of Historical Research was created, comprising the Section of United States History (formerly the Department of Historical Research), the Section of Aboriginal American History, and the Section of the History of Science. Through further reorganization, effective January 1, 1937, the Division now consists of the Section of Aboriginal American History, the Section of Post-Columbian American History, and the Section of the History of Science.

Again, in 1935, a Division of Animal Biology was established representing the administrative merging of the Department of Embryology at Baltimore, Maryland, the Department of Genetics with its Station for Experimental Evolution and its Eugenics Record Office at Cold Spring Harbor, Long Island, the Nutrition Laboratory at Boston, Massachusetts, and the Marine Biological Laboratory at the Tortugas Islands, Gulf of Mexico.

The setting up of such continuing committees as the Advisory Committee in Seismology, the Committee on Study of the Surface Features of the Moon, and the Committee on Coordination of Cosmic-Ray Investigations (committees consisting of members of the Carnegie Institution staff and of other agencies) represents still another response to President Merriam's belief in the effectiveness of group action in the conduct of research.

As to researches exemplifying international cooperation, conducted jointly by Institution workers and representatives of other agencies, numerous examples are to be found in the reports of department heads as recorded in the Year Books of the Institution.

Grouping of personnel for cooperative and collective research, then, is a conspicuously distinctive feature of the present-day practice of Carnegie Institution. Without doubt, the achievements of the Institution during the recent years of its activity have been due in no small measure to judicious employment of this method of attack.

In his efforts to bring about administrative reorganization advantageous to cooperative endeavor, President Merriam had the un-questioned support of the Honorable Elihu Root, who from the founding of Carnegie Institution to his death in 1937 was a member of the Board of Trustees and its chairman during the last twenty-four years of his life. Although Mr. Root approved of the rule which Mr. Carnegie used with great success in his business—"Find the exceptional man and give him a free hand"—nevertheless, he believed that the effective power of even exceptional men may be increased by organization, as the following excerpts from his remarks on the subject will show.

> Science has been arranging, classifying, methodizing, simplifying everything except itself. It has made possible the tremendous modern development of the power of organization which has so multiplied the effective power of human effort as to make the differences from the past seem to be of kind rather than of degree. It has organized itself very imperfectly. Scientific men are only recently realizing that the principles which apply to success on a large scale in transportation and manufacture and general staff work apply to them, that the difference between a mob and an army does not depend upon occupation or purpose but upon human nature; that the effective power of a great number of scientific men may be increased by organization just as the effective power of a great number of laborers may be increased by military discipline. . . .
>
> This attitude follows naturally from the demand of true scientific work for individual concentration and isolation. The sequence, however, is not necessary or laudable. Your isolated and concentrated scientist must know what has gone before, or he will waste his life in doing what has already been done, or in repeating past failures. He must know something about what his contemporaries are trying to do, or he will waste his life in duplicating effort. The history of science is so vast and contemporary effort is so active that if he undertakes to acquire this knowledge by himself alone his life is largely wasted in doing that; his initiative and creative power are gone before he is ready to use them. Occasionally a man appears who has the instinct to reject the negligible. A very great mind goes directly to the decisive fact, the determining symptom, and can afford not to burden itself with a great mass of unimportant facts; but there are few such minds even among those capable of real scientific work. All other minds need to be guided away from the useless and towards the useful. That can be done only by the application of scientific method to science itself through the purely scientific process of organizing effort.

The grouping of specialists for cooperative effort implies the outlining of a carefully considered program of action else confusion and waste of time and funds will ensue.

Until comparatively recent years discovery was largely fortuitous, a matter of happy, haphazard chance. In instances, crude experiments were performed which were followed by fortunate guesses later to be proved correct by more rigorous methods. In other instances, effort was made to apply the method of experiment outlined by Francis Bacon, with whom the era of modern science may be said to have begun. Bacon would have the investigator record all available facts, of whatever kind; make all possible observations, of whatever sort; and perform all feasible experiments, unrelated though they might be. Then, by collecting and tabulating the results he expected that connections between phenomena would automatically appear, leading to formulation of general laws. But even under the Baconian method, the researches of earlier periods, only in a very limited sense, can be called planned researches.

Not until the coming of agencies definitely founded and financed for the prosecution of research, with staffs of specialists on full-time salary, and with heavy investments of capital in buildings and equipment, did research take on aspects of a business enterprise, chief feature of which being the adoption of a planned program of work, comprehensive and forward looking in general outline, detailed and definite in respect to daily schedule.

But a program of research must be kept flexible, sensitive to quick change when new and upsetting facts come to light. Even the most gifted investigator is obliged to feel his way, gropingly, into the unknown, and, as President Merriam has said: "However narrowly confined the program for such an investigation may seem to be, attainment of the goal generally involves many shifts of procedure."

And this surely, this shift of procedure, has been the experience of the investigators who are engaged in study of photosynthesis, as a review of their work will show. A plotted chart of their course would reveal no straight line running between starting point and goal; rather the pathway is a twisted, tortuous, circuitous one, one which doubles many times upon itself; while the goal, the ultimate goal—the complete explanation of nature's remarkable food factory—despite twenty-eight faithful and arduous years, has not yet been even fairly approached.

Advance is necessarily slow and characterized by small gains won through exercise of infinite care and patience. Observation and reflection give rise to questions, questions suggest possible answers, the assumed answers, one by one, are put to rigorous experimental test. Not until the evidence thus derived—evidence which, if valid, can be obtained and checked by any competent person—completely sustains an assumption, are the investigators ready to declare that another step toward the goal has actually been taken.

This feature of research procedure has been happily described by Dr. Spoehr in the third of the Elihu Root series of Institution lectures on the influence of science and research on current thought. Commenting upon the immensity of the problems confronting science and upon the way in which their solution has been sought, he said:

> Science has attacked these problems in a stepwise manner. It does not hope to arrive at ultimate truth by one master move or by a single brilliant idea. It has long realized that the formulation of concepts is a matter of evolutionary development and it has planned its attack accordingly. One small and carefully planned advance has secured a position from which another advance could be made, and so on, step by step, the development has been secure and remarkably rapid.

Consideration of the experience of Dr. Spoehr and his associates suggests, furthermore, that in a research dealing with fundamental problems, such as is that of photosynthesis, each advance brings new questions, raises new and deeper difficulties. With singularly penetrating observation Emerson expressed this with exactitude when he wrote: "Nature will not be Buddhist: she resents generalizing, and insults the philosopher in every moment with a million of fresh particulars." Moreover, in the same vein Albert Einstein has recently declared: "Science is not and never will be a closed book." If these statements be true, then it follows that there cannot be now or ever any part of nature that will be fully revealed, nor can there be now in nature, or ever, a final explanation of any really fundamental matter. In consequence, to liken the objective of research on basic problems to a traveler's goal, a definite place that can be arrived at, is to employ similitude ineptly; rather, the truer representation of the scientist's experience in prosecution of research is that afforded by the pilgrim in William Drummond's "Flower of Sion":

> Ah! as a pilgrim who the Alps doth pass,
> Or Atlas' temples crown'd with winter's glass,
> The airy Caucasus, the Apennine,
> Pyrenees' clifts where sun doth never shine,
> When he some heaps of hills hath overwent,

> Begins to think on rest, his journey spent,
> Till, mounting some tall mountain, he do find
> More heights before him than he left behind.

Further consideration of the activities of the photosynthetic workers gives clear indication that achievement in the particular sector occupied by the group is to be credited in part, at least, to the advancing and supporting front of science. Although much sound exploratory work had already been done when the group of scientists began their studies, as has been stated, it was necessary to reexamine theories then dominant with the result that some were accepted, some were modified, and others were discarded altogether.

As to contributions from other fields, atomic physics provided concepts that aided powerfully in study of the molecular structure of photosynthetic products. The invention of the spectroscope and the development of the astonishing technique in its use were important factors in evolving an extremely sensitive method of separating and isolating plant pigments. Except for the advance which had taken place in organic chemistry, many experiments in course of the investigation leading to important conclusions could not have been made. Study of metabolism in man and animals produced information which has been of definite aid in the investigation of the phenomenon of respiration in the photosynthetic process. And, to suggest but one other illustration, the development of instruments and techniques for making measurements more precise, particularly in dealing with "the world of neglected dimensions," a development that is one of the marvels of modern science, has brought definite gains in study of photosynthesis that otherwise would have been impossible.

And so, scarcely a major advance in any branch of scientific investigation has taken place, the results of which have not been levied upon by the workers in photosynthesis. The progress made in this investigation, therefore, is by no means to be credited to a single mind or to a single school, as Dr. Spoehr has pointed out. In fact, in this matter he likens the enterprise that he and his associates are engaged upon to a structure in the building of which a thousand workers, representing many nations, have labored to quarry and fashion the stones and to assemble and fit them into place.

In all essential respects, the story of the progress of this research in photosynthesis is the story of the progress made in every other sustained investigation that the Institution has entered upon. The Year Book reports covering the work of each of the major departments and divisions of the Institution's organization contain numerous illustrations of how present-day research is conducted, equally typical and illuminating.

In every instance, it will be found that the investigation starts with ground already partly plowed by others; that its progress is helped through the direct and indirect cooperation of many investigators, representing many fields and agencies; that it is guided by a leader and executed by specialists who become so saturated with the subject that pertinent questions and plausible answers arise spontaneously and intuitively; that the practical consequences of the assumed answers are defined and then tested by observation or by controlled experimentation; and that when disharmonies appear, a new hypothesis is framed, and so on until one is found that is compatible with all the facts that have come to light. Sooner or later, if the course be followed persistently, a hypothesis will be suggested which adequately integrates all the known facts that are relevant—thus a theory is born. The theory, in turn, may serve to coordinate and simplify knowledge for many years, only to be modified or superseded entirely as new tools for better measurement subsequently produce fresh, disrupting facts.

In short, diverse as are the problems upon which the various Institution groups are engaged and diverse as are the uses to be made of the results obtained, the workers are all embarked upon a common quest: to discover significant facts by hypothetical anticipation coupled with planned and discriminating effort and to arrange them into cause-and-effect relationships or into such other relational groupings as will suggest postulates—a first step on the road leading to demonstration of order in nature and the formulation of laws expressing it. And in this respect the staff of Carnegie Institution is typical of research workers in pure science generally, wherever they may be situated or under whatever auspices they may be working.

IV

I turn now to consideration of the system of ready communication which has developed among the research workers of the world whereby information about the progress of a given investigation, wherever conducted, is quickly circulated.

Meetings of scientists (sectional, national, and international) are held frequently, providing opportunity for personal contacts. Papers of importance are presented and discussed at these meetings whereupon, in whole or in abstract, they are published in the proceedings of the meetings or in the journals of the various sciences and thereby made readily accessible. Usually, when an investigation has successfully passed a definite node in its progress, announcement is made through the columns of some journal representing the field in which the research lies. Again, at intervals, a survey and summary of what has been accomplished is prepared and the report published as a brochure or, indeed, if more pretentious, as a monographic volume or a series of volumes.

In such manner, accounts of the progress of scientific inquiry are brought promptly to the attention of the research workers of every nation. They are also deposited in the principal state and institutional libraries of the world, there to be consulted by all who desire the opportunity. Thus, through this system of interchange, the specialists in the various fields of investigation are kept informed of the activities of all the other workers, both at home and abroad—kept informed of the methods employed, of the tentative conclusions reached, of the demonstrated results announced. Without doubt development of this system of intercommunication among scientists has been a factor of the greatest importance in the phenomenal advance which the research of the present century has experienced.

Recognition of indebtedness to such interchange has been given by no less a person than Albert Einstein, whose contribution to present-day scientific thought places him in the rank of the greatest scientists of all time.

"Many times a day," he has said, "I realize how much my own outer and inner life is built upon the labors of my fellow-men, both living and dead, and how earnestly I must exert myself in order to give in return as much as I have received. My peace of mind is often troubled by the depressing sense that I have borrowed too heavily from the work of other men."

For the most part these published reports are designed for the use of colleagues in the same or in cognate fields; they are not intended for the general reader who has had no scientific training. The concepts described and discussed, in themselves, are unfamiliar; moreover, the vocabulary made use of is one so highly specialized, indeed, that workers in branches of research other than that in which the report originates may not themselves understand the terms employed. The language in which the papers and reports are written is one that has evolved in response to effort to secure brevity in presentation and explicitness in the transmission of meaning. New terms are invented and introduced and special and restricted meanings are given to old and familiar terms. Moreover, the reports assume that the reader is in possession of background facts and theories which only the specialist can be expected to have. In consequence, nearly the whole of the subject matter of these research reports is unintelligible to all who lack special preparation.

Yet the intelligent layman is keenly curious about the physical world in which he lives. The mystery of the universe is an unceasing challenge to his imagination and he longs to penetrate it. Certain of the great scientific generalizations have caught his attention and he is eager for more. He recognizes that science has profoundly affected the manner of his living and expects it to continue to produce far-reaching effects. Moreover, he knows somewhat of the attitude of science toward its problems and the methods employed in the effort to find solution and thinks, perhaps vaguely, that the habit of mind which submits every idea to rigid test may be useful in solving the social problems that confront us. These and other reasons impel him to ask the scientists to share their knowledge with him.

It must be clear, however, from what has been said about research programs, about the methods employed by research workers in securing their advances, and about the nature of the accounts describing their work which are circulated among fellow investigators, that the layman who looks only to the progress reports of the scientists would fare badly in his quest for new knowledge. Moreover, even though the general reader could understand the language in which the reports are couched, the immense quantity which the presses are turning out would overwhelm him completely; besides, the reports, for the most part, treat of details which would be without meaning to him and from which, in consequence, he would derive neither pleasure nor profit.

Take for example the reports of the staff published by the Carnegie Institution—the thirty-seven Year Books of sixteen thousand pages, and the seven hundred and twenty-five monographic volumes of more than two hundred thousand pages—also, the eleven thousand five hundred Institution papers (estimated), aggregating one hundred fifty thousand pages (estimated), which have been published by other agencies. This body of material, representing the productivity of one agency alone, stands as a rich mine of information respecting the activities of Carnegie Institution; moreover, in so far as the Institution has covered or paralleled the course taken by research generally, the material comprises a veritable history of the development of scientific research during the present century. As such, it is inconceivable that a time will ever come when the records thus published and preserved will not be consulted by the research workers of the world; yet little of it is easy reading, even for the staff itself, and much of it is frankly difficult for anyone. The layman who wishes to learn about the progress of research, therefore, as set forth in such reports, will find direct approach to be futile; his only alternative is to betake himself to the interpreters.

It is essential to the continuing progress of research that such reports be prepared and that they be promptly published and widely circulated among scientists. It is highly important, also, that information about the significant facts and relationships disclosed by the researches represented be put in form such that it may be incorporated quickly into that great body of common knowledge which is one of the constituent parts of our cultural background and of our social heritage.

The social psychologists tell us that groups possess a mind and consciousness that are tangible entities, quite as much as, indeed, are the mind and consciousness of the individuals comprising the groups. The mind of the individual, they say, is a residual storehouse of experiences, knowledge, thoughts, emotions, which are compressed into intuitions and feelings and formulated into principles which in turn direct and modify all future experience. In like manner, it is held, groups are also residual storehouses of experiences, feelings, emotions, information, common to their members, which, as with individuals, direct and modify the future experience of the groups. As time passes, bit by bit these accumulate into bodies of common feelings, general desires, accepted appreciations, a recognized moral code, a public opinion, a general will—all of which being connoted by the terms, "the social mind," "the mass mind."

Among the furnishings of the mass mind is that aggregate of knowledge which is common to the individuals who constitute the larger social groups such as the community and the nation. Scientific inquiry is one of the important sources of this knowledge and, more and more, the mass mind is looking to it for larger and fresher contribution of its substance. Under leadership of President Merriam, Carnegie Institution is now definitely committed to the view that this demand is legitimate and that a need is indicated that should be satisfied.

Repeatedly President Merriam, in the statements explaining the work of the Institution as an organic whole which accompany the reports of current investigations to be found in each of the Institution's Year Books, returns to discussion of the ultimate values of scientific research and to the view that a given investigation has not fully accomplished its purpose until its principal findings have been stated so simply and clearly that they command easy entrance into that totality of knowledge which passes current among educated readers who are without special training. It will be instructive to consider certain of these statements.

The 1934 Year Book, for example, contains one of President Merriam's most penetrating comments on the deeper values to be derived from scientific research. It follows in full:

> The vigorous discussion by scientists and philosophers concerning worth or significance of research contributions, has centered in considerable measure around questions which have to do with maintenance of life, with amelioration of adverse physical conditions, and with factors assumed to affect the economic or social balance.
>
> As yet comparatively little attention has been given to the idea that science may have made its most noteworthy contribution through influences which aid in determining attitudes of mind and objectives. In other words, we have been concerned more largely with discussion of the extent to which science affects our environmental conditions than with the possibility that it helps to give us new points of view and a bettered attitude toward life. The great significance of this difference becomes apparent when we consider that influences determining point of view and attitude furnish major sources of human initiative, and are among the most important guiding elements in life.
>
> Science has made vast contribution toward betterment of living conditions through ready production both of necessities and luxuries, and by freeing us from

drudgery. But abolishing poverty in the material sense might not prevent poverty of mind and soul. Increased freedom from the labor entailed in maintenance of life may still leave us enchained in spirit. Wealth in the sense of natural resources or accumulated results of human toil, or even expressed through increase of capacity for work might lead only to degradation. There may be wickedness and bitterness and infinite discontent with riches and great power. Whether the world really becomes a better place in which to live depends in large part upon our attitude toward life and our ideals. If science and research help to improve these controlling factors in life, they attain the supreme service. If bettered point of view can arise only from other sources, then science and civilization must wait until some way is found by which there may be exerted the power necessary to bring about the change desired.

The influence of science upon our point of view and ideals may be thought by some to be negligible, since the attitude involved will be assumed to depend upon spiritual values, and research is looked upon as concerned only with cold facts and logic. It is, however, important to realize that by definition science represents the seeking for truth, whether it relate to the elements of physics, chemistry, history, or human conduct. The use of reason is only the logical extension of truth. Facing the facts by the method of science is to strip away untruth, dishonesty, self-deception. This may be by changing alchemy to chemistry, the rabbit's foot type of healing to scientific medicine, wishful thinking to factual economics, or self-centered beliefs to constructive religion.

It must also be recognized that a view over the vast range of things expressed through science as in astronomy, earth building, biology, or the story of life through the ages, gives a greatly increased appreciation of law and unity in the world. Such a view includes all history and our relation to it. It presents a new outlook over the universe, with a clearer vision of man's place in the scheme of things, a better opportunity for appreciation of what life represents, and a changed attitude toward its problems. Seen in this light, science should aid in the forming of basic beliefs and philosophy, and even religion may use it as material with which to build. Science gives reasons why every man should have a philosophy and admit it, and at least an appreciation of what religion may signify.

In considering the relation of science to those fields of thought which are generally deemed most clearly to express human interests, it is desirable to suggest that the difference in attitude among these subjects is not necessarily as great as is sometimes assumed. Science finds need for common ground with philosophy, art, and religion in the work of developing a clearer, broader, and deeper vision of the world of things and of people about us.

As seen by science, the universe is a vaster and more orderly, more dependable place in which to live than was once recognized. At the same time the scientist may appreciate more clearly today than at any previous stage that he does not really fathom nature in essence, or power, or ultimate meaning. With the advances made by modern science the so-called material universe does not grow more definitely material, or at least one may say it is still beyond our full understanding. And for these reasons we need close relation among the various points of view which we may take. The interests of science, art, philosophy, and religion

> must be joined if their human value is to be most fully realized. Each may stand alone as an abstract or non-human value, but when human interests are touched they must come into intimate, mutually supporting relation.
>
> Science we see as a powerful educator because it turns attention to real things, and not to substitutes. One can not dismiss the idea that these expressions of reality and law relate in large part to things other than mere food and maintenance of life. We need wide acceptance of the attitude of mind illustrated by science in searching for facts upon which to base judgment.

Again, in his 1937 Year Book report President Merriam discusses the influence upon society, in respect to the solution of the so-called social problems, that science may be expected to exert. On this subject he says, in part:

> The much-desired cooperation among students of human affairs through joint effort of workers in the fields of the natural and social sciences, such as economics and government, has seemed less easy of realization than was perhaps expected. There can be no doubt that certain of the attitudes and methods of science appear difficult to use in application to human relations, and that the methods applied so directly in the natural sciences often seem inadequate in research upon our vastly complicated human questions. There are, however, many aspects of research procedure developed in the natural sciences which will have increasing value in application, either directly or indirectly, through various phases of the social sciences.
>
> There is importance in the idea that while in certain aspects of natural science, as in physics and chemistry, one seems to deal with immutable or fixed laws, the factors presented for consideration in social science broadly, or in economics, shift so frequently that the methods of natural science can be applied only with difficulty. It is, however, well to remember that problems in biology also show us a situation in which changing conditions, and apparently changing factors, present one of the recognized features in any investigation. The larger history of evolution in life-forms gives one of the best illustrations of continuous change, and yet these changes seem to depend upon fundamental laws which may continue in operation much as in physics or chemistry while new factors come and others go. The general principle of evolution rests upon conservation of certain features which might be called fundamental, and continuous shift in others leading to production of new situations or features which may bring about what we call creation.
>
> In the rapid shift of human affairs, conditions, opportunities, and obligations change so rapidly that we may tend to view the laws of conduct as themselves changing fundamentally. So we seem almost to find in international diplomacy a new statement of human obligations, and a new interpretation of principles which have had general recognition among individuals over the world. And yet, we may have faith that things which are looked upon as representing the principles of right and honesty will in the end prevail. Perhaps neglect to appreciate this point of view is responsible for some of the difficulty in the world at the present time.

It is also to be borne in mind that we must not only recognize the principle of change or evolution in whatever relates to life, but we must give recognition to the principle of constructive or creative activity developing in the peculiar types of creatures known as human beings. But creative activity does not rest upon a breaking down of law. It is a natural expression of certain laws through means which make evolution and the development of new elements possible.

Orderly movement in social evolution probably depends in considerable measure upon our ability to determine the balance which should obtain among certain factors which represent immutable laws, others which we find leading to evolutionary development, and still other features in human life which have to do with constructive or creative activity. Such matters may become of great importance in the future.

One aspect of this problem, which seems often to be misunderstood, has to do with the influence of science upon society in the sense of utilization of facts merely as facts, and not as parts of a wider scheme of knowledge. It is possible to bring extreme confusion into human affairs through the use of facts. This may be illustrated when on two sides of a legal case opposing talent uses scientific information, but in each case covering only a part of the evidence. Facts are used, but not all of the facts, and such situations make controversy possible. It has been said that truth differs from fact in that truth represents all of the facts, or, as is said commonly, all of the truth, so set up as to give a clear picture of what exists. It has also been said that honesty is the expression of fact or of the truth as involved in discussion of human relations.

Concerning the less obvious values of science, I am tempted to quote the paragraphs in the President's report of 1937 which deal with science and appreciation of nature. These follow:

One illustration of a way in which science is held to affect present-day thought is found in the suggestion that research tends to diminish the interest of mankind in the world of nature about us, which is the principal field of scientific activity. This is in spite of the fact that we are practically embedded in nature, and that much of man's philosophy of life in past centuries has based itself upon observations concerning the character of the world in which we live and our relation to it.

The argument in favor of the view that there is diminishing interest in nature on the part of modern man rests in considerable part upon the idea that present-day tendencies look toward shift of human appreciation to results of mechanization in the world, or in the life of man, by reason of advances in science. It has been thought that jazz, the radio, the automobile, and many other contributions of present-day thinking have turned interest away from the elements of nature and that, considered broadly, mankind of today transfers its attention to the varied, and often mechanized, activities of man, rather than to direct it toward the tremendous range of evidences of creation in the world about us.

If it be true that the coming of science has diminished interest in the values of nature, serious harm has been done. Much of our reverence and religion of the past had its origin in what we knew of the world of nature, and much of the soothing influence exerted upon man through his environment has been derived from the appeal of the natural world and love of its beauty. It is, therefore, a matter of great human importance to consider whether the influence of science in this respect is really detrimental.

Whatever argument is presented on one side or another of this question will almost inevitably become complicated, but it is important at the moment to call attention to the fact that with the extension of our knowledge of the universe through science in all directions, by the telescope into the infinite spaces inhabited by spiral nebulae, by the microscope into the infinite littlenesses of the world of bacteria and molecules, or by the student of history through the vast reaches of time, we have come to see the world as almost infinitely elaborated, and yet having all of its parts intimately related. Therefore we think with some assurance on the idea of unity of nature. At the same time we have seen movement of the world through the ages develop that aspect of being which we call evolution, producing further developments, not only more complicated, but which we consider more advanced than those of preceding periods. So, seen in the light of the broader reach of science, the universe about us becomes more complex and more wonderful, and yet more dependable, more clearly creative, and more beautiful. And question may well be raised whether the influence of science through interpretation of the world in which we live does not tend to advance or develop our interest in nature and appreciation of it more rapidly than the elements of mechanization tend to turn our attention away from nature and largely toward the present-day works of man.

And, finally, President Merriam's thought respecting the importance of making research results understandable both by workers in fields other than that in which the research lies and by the non-technically trained masses is outlined in the following excerpt from the 1926 Year Book report:

Continuing study has been given in the past year to understanding of means by which the Institution can most advantageously make the results of its work available to all scientific workers and to the public. While there can be no doubt that the making of accurate and clear record of all data secured is our first duty, and is the basis for future work and application, it becomes increasingly important to obtain such statement of conclusions reached as will make them available in their simplest form, and in their clearest relationship to other knowledge, for use of specialists in all fields of science and for the public.

In this age of narrowing objectives, it is possible to make real contribution to knowledge by following carefully any one of a great number of paths leading toward the field of the unknown. This specialization of the present day can develop a state of affairs in which the worker, extending his researches along a narrow lane, may largely lose sight of the major field to which his work is related. He may also, and perhaps properly, become so much concerned with his own investigations that he finds it difficult to express his results in the terms of the larger problems.

The scientific literature of the world could easily become so technical in its various departments that advantages of special researches would be in large measure lost to workers in other fields. Such a situation would not merely deprive the world in some measure of the advantages of specific researches, but it would prevent the investigator from obtaining for his own use data required from other fields. It, therefore, becomes extremely desirable that, in addition to precise records concerning special investigations, there be made available such data as will present the results in easily interpretable form to workers in other fields of science, as also to the public. There is reason to believe that the rapid march of science would be definitely slowed in the immediate future were it not for the attempt by clear statement to bring each item of knowledge into its place in the larger scheme of information.

Recognizing that efficiency of research must not be sacrificed by reason of attempt merely to popularize results, continuing study has been put upon the problem of how we can best bring contributions from Institution activities into position where they may have the largest value to the scientific and general public. In carrying out this idea the Institution exercises in some measure an educational function, in that it attempts the transmission of information to others in such form that it may find its largest use.

V

In response to President Merriam's promptings much study has been devoted during the past dozen years to question of how information about the significant things of the Institution's researches can best be made serviceable to those agencies which are engaged in the dissemination of knowledge. Obviously this matter touches vitally the educational instrumentalities which society has developed whereby ideas, both old and new, are kept circulating among its members in readiness to be converted to use by each according to his inclination—instrumentalities, by means of which the individuals constituting the social body, like the tissue cells of the human organism, are enveloped in a nutrient-bearing stream. Among important agencies thus serving the social order are: the school, the press, the exhibition hall, the lecture platform, the radio, and the motion picture film, all of which Carnegie Institution has made use of more or less extensively.

The school in one of its aspects is an institution created by society for the purpose of training its youth to social efficiency. Looked at in another aspect, it is that agency upon which is laid a heavy share of responsibility for transmitting the accumulated store of learning from generation to generation.

Largely because of the school, though not entirely, each generation, in respect to learning, begins where the preceding one left off.

Were it necessary for every generation to start back at the same level and to work of and for itself alone, progress would be impossible. Even though perceptible gain were achieved in a given period the advantage would be lost to all who come after. As it is, however, the experiences of each generation are organized, systematized, catalogued, labeled; and through speech, the printed page, and the conscious methods of institutions created for the purpose, the cumulative achievements of the race are handed on to posterity.

In point of fact, the genesis of the school lay in recognition of need for transmitting accumulated knowledge. Although this function by no means compasses the whole of the school's activity, nevertheless, without doubt, the development and marvelous expansion of the school among civilized races has been due, in considerable degree, to the faith which people have held that through its work this task could be accomplished satisfactorily. Yet it is by no means practicable for the school to transmit the whole of knowledge; it can, however, acquaint the youth with many of the important elements of race accomplishment and show them how best to proceed if they wish to draw upon the repositories for more information.

Among the essentials of the school's organization is the teacher, and the teacher's principal tool has always been the textbook. Until recent years, the textbook, in most subjects, was the chief if not the sole source of the content of instruction. Whatever found place in it was taught; whatever lay outside was ignored. Today, however, scientific discovery is proceeding at such a rapid pace that textbooks are obsolete almost as soon as they are written.

Under sharp attack of critics, the demand for a richer content in school instruction, and the better preparation of teachers, and the need of keeping abreast of new knowledge, the textbook has been relegated to the minor rôle of serving as an outline of the subject it represents and as a compendium of important facts and generalizations useful for review purposes. To meet the situation the teacher now seeks to increase the wealth of her own mind and to clothe the dry bones of the text through use of collateral material that is rich in concrete detail and that is interestingly presented.

Moreover, in response to demand for enrichment of instruction educational administrators who are responsible for outlining courses of study are turning more and more to the sciences. Indeed, serious discussion is now on among them as to question of a twelve-year sequence of science instruction for schools below college and university rank. Proponents of such a program urge it because they believe that the person who takes such a course will be a different person from one who has not studied science. They assert that his behavior will be affected thereby in certain major ways: through the better understanding that he will gain of the forces, phenomena, processes, materials, and living things that interact to produce the world in which we live; through development of the scientific attitude of mind that will accrue to him; through the training that he will get in the scientific method of thinking. Already, in certain city school systems practical steps have been taken looking toward the initiation of such a program.

Thirteen years ago, Carnegie Institution, aware of the growing interest of schools in science instruction and of the need of teachers for material that would keep them in touch with the progress of research in science, undertook to supply schools of junior and senior high school rank with simply written articles interpreting the researches conducted by the Institution staff. These articles are intended to furnish the background setting of a given research, the setting that is usually omitted from the progress reports as supplied by the investigator. They seek also to tie the significant things of the research in with information that the teacher and her pupils may be presumed to have. Also they are profusely illustrated with the best pictures pertinent to the text that can be obtained, for it is recognized that appeal to the eye is an important aid in the pedagogic process.

The articles themselves are occasionally prepared by the investigators, sometimes they are based upon notes supplied by them, and, again, sometimes they consist of restatements of the progress reports as published in the Year Books and monographic volumes of the Institution and in various technical journals to which members of the staff contribute. But, whatever the source, before these articles are published, great care is exercised to insure accuracy of statement and to this end each is carefully checked by those who are responsible for the research represented.

Articles so prepared are distributed as separates among schools. However, a reserve stock is held back and when a sufficient number of issues have appeared to justify it the separates retained for the purpose are assembled, indexed, and bound, and thus made available to schools and libraries for reference purposes. Four such volumes, ranging over all the activities of the Institution, comprise the output to date in conduct of this service for the schools. Teachers and educational officials report extensive use of the material thus supplied.

Second only to the school as a distributor of ideas is the press—the newspapers and periodicals taken collectively. The staple commodity of the press, especially the newspaper part of it, is news. Editors define news in various ways but examination of their practice suggests that the derivation of the term, coming as it does from *novum*, a new thing, indicates where their hearts all lie. In consequence, whereas the school deals for the most part with organized knowledge that possesses some measure of permanent value, the press is occupied mainly with ephemeral things—ideas, situations, happenings—that come into being suddenly, live for a day, and vanish. To be sure, out of the welter of evanescent happenings a residual product is obtained that may be accorded place in the storehouse of racial possessions. But such possibility is not the criterion which newspaper editors employ in filling their columns.

With much of the material used by editors freshness is an essential quality and so, usually, a story is handled with extreme speed. Events occurring in the forenoon, in whatever part of the world, are reported in the papers of the afternoon or, at latest, in the morning papers of the next day. The so-called "spot news," that is, news that must be used at once—on the spot, so to speak—is of this character.

Examples of such type drawn from the field of scientific research are: the discovery of a giant star that had never before been seen; the identification of spectral lines in the sun's radiation, indicating presence of rare elements; the finding of fossil human remains in association with the remains of extinct animals, suggesting extreme age; announcement of a new glandular secretion that affects body growth; achievement of definite gains in study of abnormal cell growth, having possible bearing upon cancer cause, etc. In respect to such items editors, generally, accept simple, concise, unembellished statements of fact as spot news and publish them immediately.

On the other hand, editors have many stories which they hold in reserve for emergencies, stories that can be used as readily one day as another. With such material immediacy is not essential to the arousing and the holding of the reader's interest. The description of the contents of a freshly opened grave in the ruins of an ancient Maya city, for example, with background information pertaining to the cultural stage represented, can be read today or a month from today with equal interest. Virtually all the interpretative and expository accounts of research investigations are of this type, the type wherein question of time is not of paramount concern. Many of the illustrated articles printed in the weekly magazine sections of the great metropolitan dailies belong in such a group. Indeed, in many cases parts of the magazine section are made up and printed days and even weeks before issuance. Moreover, all the great news-gathering and distributing agencies send to their members and subscribers quantities of material of the kind; while the smaller newspapers of the country, such as the community weeklies, draw heavily upon this "time copy," as it is called.

Inasmuch as the press provides means of quickly reaching every part of the social body, Carnegie Institution has given much consideration to so shaping its material that it will fit easily into the conduits of transmission which the press has developed. For example, announcements of important discoveries made by Institution members or of significant developments in the progress of a continuing research, announcements that fall into the category of "spot news," are marked for immediate release and delivered promptly to the principal press associations and to the newspapers maintaining offices in Washington.

Again, articles, two or three columns in length, which are based upon the progress reports of the Institution staff, are sent at frequent intervals to the principal newspapers and periodicals of the English-speaking countries. Through the cooperation of news syndicating agencies and of special writers with foreign connections, they also find their way in translation to important papers and journals of European and South American countries. As a feature of this service, arrangement has been made to supply editors with suitable photographic prints if they indicate desire for such illustrative material. So, also, in this connection, briefer versions of a given article, together with illustrations in the form of "mats," are transmitted to the smaller newspapers which lack photo-engraving facilities.

Editors also have use for brief statements that are not illustrated. To meet this need a Clip Sheet Service has been organized and at frequent intervals short articles of four or five hundred words relating to the significant things in development of researches are mailed to the principal newspapers of the country in a form such that they can readily be clipped and sent to the composing room for keyboarding.

The attitude of editors generally toward the reporting of science news, it should be noted, is changing markedly for the better. Time was when their interest in science was limited to that which was novel and sensational. Neither scientific research nor the scientists themselves were treated seriously. Young reporters who were assigned to cover meetings and conferences of specialists and, indeed, to interview men of profound learning, with great intellectual achievements to their credit, asked questions so puerile as to be laughable. Many scientists of note have had the irritating experience of finding themselves saying in published interviews utterly absurd things about their own work and that of others. Moreover, many editors seemed to think that stories pertaining to scientific matters should be treated jocosely and even flippantly. It is not to be wondered at, in consequence, that the natural aloofness of the scientists, engrossed as they are in their investigations, was augmented even to the extent of positive refusal, in many cases, to cooperate with the press.

Happily the opinion of editors respecting these matters has altered. Having long realized that treatment of finance, of art, of music, of drama, of sports, calls for the services of specialists in these respective fields, they are now beginning to accept the view that it is also necessary to have persons on their staffs capable of dealing intelligently with scientific subjects. In consequence, the great press associations, the news syndicating agencies, and the leading dailies of the country are including science editors in their personnel and holding them responsible for all news stories and editorials touching the field of science which are transmitted or published. Definite improvement in the handling of science news has resulted and this is arousing the confidence of the scientists, who, in turn, are becoming more willing than ever before to aid the press in the endeavor to bring the important things in research within the grasp of the general reader.

Magazine and newspaper articles interpreting the work of research investigators and books about science, however vividly written or however well illustrated, cannot possibly be as vivid and as vital as the objects of knowledge themselves. A Chinese proverb puts the matter succinctly: "Hearing a hundred times is not equal to seeing once." It was the thought embodied in this proverb that dominated the work of Louis Agassiz, one of America's greatest teachers of science. In writing about him, the late David Starr Jordan, himself one of Agassiz's most distinguished students, said that this great biologist discouraged the use of books in teaching, insisting instead that it is more profitable to give contact than information, and saying in explanation: "If you study nature in books, when you go out-of-doors you cannot find her."

Although it is of course impossible to make the educational process completely objective, nevertheless, it should be realized that there is a distinction between knowledge and the objects of knowledge and that in acquiring the former it is good pedagogical practice to go as far as facilities will permit in putting students in direct touch with the latter. Believing that such procedure is especially helpful in the effort to make clear methods employed in research and the results obtained therefrom, Carnegie Institution presents each year an exhibition based upon the research activities of its staff; thus opportunity is afforded for contact with some of the great realities of science. An editorial writer on one of the Washington dailies caught the purpose and spirit of these exhibitions when he wrote: "The public is invited to the Carnegie Institution each year not to witness a show but to look upon realities—the most fundamental realities of creation."

The exhibition is held in the Administration Building of the Institution at Washington on the occasion of the annual meeting of the Board of Trustees. It is prepared by the various members of the staff, working under the general direction of a committee appointed from year to year by the President. After it has been viewed by the Trustees and their guests it is opened to the public and for three days and evenings anyone who wishes to may visit it.

The individual exhibits are set up by a selected group of investigators, the personnel of which changes every year. In preparing his exhibit each exhibitor seeks to organize and present it so that a clear idea is conveyed of some significant phase of the investigation upon which he at the time is engaged. An important factor in the success of these exhibitions lies in the fact that the investigator himself not only prepares his exhibit but is present throughout ready to explain it and to answer questions about it which the observer is prompted to ask. During the progress of the exhibition a program of short lectures is provided. These lectures are given by exhibitors and are designed to afford background explanations of the several exhibits. Radio broadcasts relating to important developments in the work of the Institution suggested by the exhibition have also been given during the period.

Apart from the opportunity which the exhibition affords for acquainting the people who attend it with the work of the Institution, one of the most consequential results lies in the favorable occasion which it provides for members of the staff to meet the Trustees of the Institution, to establish profitable contacts among themselves, and to develop a sense of Institutional cohesion.

The geographical separation of groups of Institution workers is great, owing to the fact that, in general, they are placed where their work can be conducted to best advantage. Thus, the Department of Meridian Astrometry is located at Albany; the Nutrition Laboratory at Boston; the Division of Historical Research at Cambridge, with sections working at Washington and in Yucatan and Guatemala; the Department of Genetics on Long Island; the Department of Embryology at Baltimore; the Geophysical Laboratory and the Department of Terrestrial Magnetism at Washington, with observatories of the latter in Peru and Australia; the Mount Wilson Observatory at Pasadena; the Division of Plant Biology at Stanford University, with sections at Tucson, at Santa Barbara, at Carmel, at Berkeley, and at Pikes Peak; and the Marine Biological Laboratory at Tortugas.

This physical separation of the unit groups of Institution workers would naturally lead quickly to a feeling of disconnection and heterogeneity with consequent loss of Institutional efficiency if nothing were done to overcome the tendency. The period of the annual meeting, however, with its accompanying conferences, lectures, and exhibits, provides opportunity for personal contacts, for discussion of common problems, for consideration of the larger Institution policies and program, and for stimulating interviews with the President, that go far toward counterbalancing the disadvantages of widely scattered locations.

The effectiveness of such activities, however, has fallen short of possibilities because of the inadequate facilities of the Institution's Administration Building, which was not designed for such a broadened purpose. The extension just erected, an extension which doubles the capacity of the original building, has been carefully planned to provide ample space not only for the activities associated with the annual meeting but also for housing a permanent exhibition representing results of the research activities of the Institution, for adequate conference rooms to be used throughout the year, for an auditorium of suitable capacity where lectures can be given under favorable conditions, and for office rooms to care for an expanded personnel devoted to the development of public relations.

So, too, study is being given the question of how the several departments of the Institution's organization, within reasonable limits, can provide opportunity at their respective headquarters for helpful contact with the general public.

The Institution, for example, has recently erected on Mount Wilson near Pasadena, California, an auditorium and exhibits building for the accommodation of those who desire to visit the Observatory and to learn somewhat about how astronomers conduct their investigations. One evening each week is devoted to the demonstration of the work that is constantly in progress at that point. A member of the staff gives a short descriptive talk to those assembled in the auditorium after which the group visits the 60-inch telescope near by, where each is given opportunity of looking into it and of viewing certain of the celestial objects. Thus, effort is being made to give those who are interested opportunity of gaining a measure of first hand knowledge of the work done at a great astronomical observatory. That there is need for such accommodations is witnessed by the fact that during 1936, before present facilities were installed, more than 50,000 persons visited the Observatory during the hours when it was open to the public.

In concluding this discussion of the means employed by the Institution in making available the results of its researches, mention should be made of two series of public lectures which are given at the Administration Building in Washington. One consists of general lectures by staff members on their work, delivered during the winter and spring months. These are simple but authoritative statements, illustrated by lantern slides or motion pictures, and made by leaders in the various fields of research touched by the Institution. The lectures are published and reprints made available for wide distribution.

A second series, known as the Elihu Root lectures, is likewise given in Washington; the lecturers, however, are not limited to Institution personnel. The series is devoted to discussion of the influence of science and research on current thought. The best thinkers of the present day, wherever situated, are invited, as opportunity presents, to take place on the roster of speakers. As with the series first mentioned, publication and general distribution are provided.

In addition to these established lectures, President Merriam, the departmental heads, and members of the staff of investigators are frequently called upon to address learned societies and citizen groups—President Merriam, upon the philosophical aspects of research, the place of science in the social order, and its relationship to other disciplines in the cultural equipment of the race; departmental heads and research workers, upon the details of their investigational activities. And here again provision is made for publication and widespread dissemination.

In the foregoing statement the means are outlined which have been developed by Carnegie Institution for contributing, on the one hand, to the needs of investigators who have specialized in particular regions of investigation and, on the other, to satisfying the desire of well-informed people, both technical and nontechnical, for the generalizations of research and for information about the social implications of these generalizations.

The first of these objectives is the traditional one, the one that was uppermost in the thought of those who established the Institution and shaped its policies during the formative years. So viewed, research was pursued primarily for its own sake. The immense accumulation of facts that resulted is stored in monographic publications, which, along with those of other agencies, have been likened, in respect to availability, to an enormous safety deposit vault having many boxes, large and small. Though the keys to the boxes are held by many persons, no one possesses the keys to more than a few.

The second objective in the publishing activities of the Institution, that of unlocking the boxes and making the contents accessible to all alike, is, obviously, an educational one. It calls for restatement, for discriminating selection, for integration on different levels, for generalization, for indication of significant relationships, for suggestion as to use in the affairs of life, for continuous revision of the old in the light of the new. Such activity is constructive, creative, dynamic—as befits a changing world.

Thus Carnegie Institution now is not only engaged in search for new knowledge but also in effort to make its discoveries broadly available. Its attitude in the matter, described in detail in this paper, can best be summarized by quoting a paragraph from one of President Merriam's addresses to the Board of Trustees:

> When we come to see the various aspects of our problem, and to full realization of our responsibility to the public, we shall have reached a situation in which we will see our duty, first, as an agency dedicated to active work in securing knowledge regarding fundamental problems, and, second, as an institution responsible to the whole prople for giving them in the most fully available form the knowledge that arises from our results.

TRANSMUTING SCIENCE INTO CONSERVATION

NEWTON BISHOP DRURY

Secretary, Save-The-Redwoods League
Research Associate, The Carnegie Institution of Washington

Once during a discussion in which the name of Dr. John C. Merriam was mentioned, a prominent highway engineer, not unkindly, referred to him as "an impossible idealist." A phrase like this stays in the memory; particularly when, as in the present case, subsequent developments have proved it contrary to the facts. Harassed by the exigencies of affairs in which not principle but expediency often determined policy and action, this engineer doubtless spoke with a touch of envy. In reality, he paid tribute to an attitude of mind which has made Dr. Merriam in many fields a contributor to the constructive thought of his generation.

No movement in which he has been engaged shows better his idealism than the conservation program of the Save-the-Redwoods League. And none by the accumulated results of twenty years more fully proves that, when sustained by courage and clear thinking, idealism brings results—is very far from futile or "impossible."

For the concrete evidence is at hand. The Save-the-Redwoods League "started from scratch." When in 1917 John C. Merriam made his now historic trip with Madison Grant and Henry Fairfield Osborn by automobile over wagon roads that have since developed into the celebrated "Redwood Highway," not one stick of timber in that region, comprising what many now consider the finest forests of the world, was publicly owned. All was in the hands of lumbering operators, and the cream of it was rapidly being destroyed. They were appalled by the devastation of forests whose ancestry extended back millions of years, with individual trees as old as the Christian era, attaining age, height, and size equaled only by the companion species of the Sierra—the *Sequoia gigantea* already largely preserved in national parks and forests. As a result of that memorable journey they realized at once that the Coast redwoods, the *Sequoia* *sempervirens*, were equally a priceless possession of the nation, and as such must be protected. They wrote the governor of California. They called upon Joseph D. Grant of San Francisco, a pioneer Californian imbued with love for his native state, and planned an organization to prevent the impending catastrophe. Many others joined. A call was sounded to the state and nation to preserve the best of this heritage before it was too late; and thus the Save-the-Redwoods League was launched.

Today, twenty years later, while much remains to be done in rounding out the program, the vision of these men has become a reality on a much larger scale than at that time they could have dreamed. In Humboldt and Del Norte Counties, where the *Sequoia sempervirens* reaches its supreme development in a setting of rich associated flora, more than thirty thousand acres of this matchless forest have been preserved in state parks representing a cost of over $6,000,000. Extending for thirty miles along the Redwood Highway, these majestic colonnades that travelers come from the ends of the earth to view, are assured of preservation for all time. But more important still, away from the beaten track, and unpenetrated by automobile roads, large units of virgin forest have been incorporated in the state reserves with all their primitive qualities undisturbed, capable of being held untouched by artificial intrusions for many generations to come.

These are in four main areas. Within the twenty-thousand-acre Humboldt Redwood Park, forty-five miles south of Eureka, the Bull Creek watershed has been preserved almost intact from ridge to ridge. At its heart the heavy stands of giant redwoods on a thousand acres of extensive flats, from South Fork of Eel River four miles west along Bull Creek, make up one of the greatest of all forest spectacles. The comparison to a mighty cathedral has been often made, yet there is no other simile to express its overwhelming effect. "Dim, lofty, pillared, peaceful, this temple which the Great Architect has been building for a score of centuries is incomparably nobler, more beautiful and more serene than any created by the hands of man."[1] Aside from the main highway, approached from the south through a series of impressive memorial groves, and from the north over the privately owned "Avenue of the Giants," which the League is now endeavoring to save, Bull Creek Flat represents the climax of a trip through the redwoods.

---

[1] Duncan McDuffie, in "Saving the Redwoods." Save-the-Redwoods League, 1924.

The Prairie Creek Redwoods, to the north of Eureka, near Orick, make up another large primitive tract, quite different in prevailing atmosphere, more somber, the dense and even stand of redwoods interspersed with spruce and fir and hemlock, with broad-leaved maples along the stream, the forest floor luxuriant in great tangles of shrubs and giant man-high ferns. Then going northward one comes to the Del Norte Coast Redwoods, seven miles south of Crescent City, where rugged seascape vies for attention with redwoods growing almost to the ocean's edge, and thickets of rhododendrons which bloom in the early summer. Comparable in impressiveness are the Mill Creek Redwoods, northernmost of the outstanding groves, where a beginning has been made upon the fourth of the League's main projects.

For under the leadership of Dr. Merriam—first as chairman of the Executive Committee, and for the past seventeen years as president—there has been developed, after long study of the redwood region, a definite program embracing these four areas as typical of the redwoods at their best.

Other matters have concerned the League, and as need arose its functions have broadened, but never during the years, having once determined upon them by exhaustive investigations, has it been diverted from these four primary objectives. In this and other respects there has been careful planning. A "balanced program" was the ideal set forth almost from the beginning: one in which account was taken of each element concerned with the use of the great resource involved in the redwood forests. Recognition of the rights and important functions of those engaged in the economic utilization of these trees in the lumber industry; regard for the judicious use of the recreational possibilities; study of the growth habits of the redwood, and its place in the evolving forest community with all its attendant life; above all, appreciation of the inspirational values, of the beauty and meaning of this work of nature, to be preserved in its most representative phases with undiminished freshness—all these considerations entered into the program whereby the League proceeded.

As the practical work of money-raising and acquiring forest lands progressed in cooperation with the state park authorities, and as step by step the reservations grew in size, it became apparent that public ownership alone did not satisfy the need for protection of these forests. There arose threats almost as great as destruction by lumbering. Protection from fire, from highway development, from inappropiate human use, involved problems for the state which have not yet been fully met, and in their solution the League is assisting.

Viewing the way in which the popularity of the redwood parks was causing them to be overrun and deteriorated, Dr. Merriam once remarked: "I sometimes think that we are undertaking the impossible task of obtaining the pleasures of solitude for all the people of America at the same time."

These problems are still being coped with. Through study, observation, and application of the best experience obtainable, emphasis is being placed upon formulating a policy for protection which recognizes the highest uses to which these properties of the people can be put. All this has contributed to the setting of a tone and standard for administration of these areas which it is hoped will have great significance for the future of the redwoods, and of other public reserves as well.

"Make a clear statement of the highest values" . . . "Establish a policy" . . . "Translate this policy into intelligent administrative action" . . . "Resist compromise with lesser ideals." These were among the utterances with which Dr. Merriam called this conservation project to its present high plane.

Realization of the importance and diversity of the program was summed up by him at an annual meeting of the Council of the League, with a statement that the stage of acquisition was only preparation for the great work of use for what had been acquired:

> The responsibilities in future activity of the League involve careful study on every aspect of use of the redwood on a plane higher than has yet been attained. Without such concentrated research and application of the results the efforts of recent years might lead only to utilization of large values on the level of the commonplace. What might be a great vision would then have been glimpsed, but not actually seen or realized.
>
> Some of my most careful thinking in recent years has been devoted to the idea that we are entering a period in which human kind, after extended sojourn in the deserts of civilization, begins to look back upon primitive nature as something of exceptional value. The elements of our surroundings, which at one time seemed to be accepted without appreciation, now show themselves possessed of tre-

> mendous values, and we begin anew to make intimate acquaintance with features of special significance to us. In reality, contacts with nature through the vast period preceding the rapid growth of civilization had a very great effect upon mankind. But it is possible that in the future the moulding influence upon human thought will exceed anything that we have known in the highest values of nature appreciation.[2]

He then went on to outline five main lines of inquiry, involving scientific study, economic application of the facts thus learned, recreational possibilities, educational effort, and æsthetic and inspirational interpretation. Only through all these could the highest value of what has been acquired be realized.

Almost from the beginning of the movement, able groups of men and women, inspired by Dr. Merriam, have been devoting earnest thought to these various studies. Much knowledge as to the significance of the redwood forests to the people has been accumulated. Interpretation of these facts has been essayed, in publications, in the work of nature guides, and notably in the exhibit at Richardson Grove, Humboldt County, where Emanuel Fritz, forester of the University of California, has used the cross section and root system of a fallen redwood, 1200 years old and twelve feet in diameter, to reveal its age in terms of historical events, the story of vicissitudes shown in its growth (such as charred fragments of bark embedded within its annular rings which clearly indicate fires occurring long before Columbus came to America), and, adding a new chapter to the knowledge of the habits of this species, a series of root systems thrown out successively anew as alluvial deposit through the centuries raised by over fourteen feet the river bench upon which this redwood had grown.

Close touch with reality, essential to a scientist, has always been emphasized by Dr. Merriam as necessary to an understanding of the meaning of great natural phenomena. Thus his impatience with the dead exhibits of museums, when they attempt to supplant the works of nature. His years of experience with the finer of the national parks bore fruit not alone in clear and uncompromising statement of principles presenting an ideal to govern the use of these masterpieces of creation. In at least two cases his patient study and planning brought the establishment of mechanisms whereby the visitor, be he casual or student, cannot fail to realize the significance of what is before him. At Yavapai Station on the brink of the Grand Canyon, means are provided for reading directly from the book of nature, and gaining a conception of the dramatic story unfolded by this "abyss in time." Likewise, at Sinnott Memorial, the beauty and meaning of Crater Lake is revealed from a point of vantage where the beholder is guided to see and understand. The same principles have been applied to the redwoods.

---

[2] John C. Merriam, "The Tasks Ahead of the Save-the-Redwoods League." Save-the-Redwoods League, 1934.

He often referred with amusement to an early episode in which a well-meaning lecturer endeavored to educate visitors to a redwood park by taking them into a small auditorium, and showing diminutive motion pictures of the very grove which, in all its impressiveness, was outside all around them. Interpretation of the redwoods is now taking quite a different course.

On the less material side, that of study and expression of intangibles such as the elements of beauty in the redwood forests, emphasis has also been placed. The best that has been said and thought on this subject has been collated. What has been done through the medium of pictorial art has been reviewed. Encouragement to further analysis and statement has been given, for Dr. Merriam holds that the difficulty of this task should be no bar to its attempt.

Through the Carnegie Corporation the initial funds have now been provided for continuing investigations in this field. These will undoubtedly follow the pattern of the Point Lobos studies, also supported by the Carnegie Corporation at the instance of Dr. Merriam, in which over a period of several years and at a cost of $15,000 specialists in many types of knowledge have appraised all values inherent in a primitive landscape of exceptional beauty, containing the sole remaining natural grove of the picturesque, wind-blown Monterey cypress, associated with many interesting plant and animal species in an undisturbed environment. The master plan and policy formulated for the management of the property with the aim of perpetuating highest values in Point Lobos Reserve, three miles south of Carmel, have already had a profound effect upon administration of the park lands held in trust by the state of California.

In another field, related mainly in manner of approach, important guidance has been given. The old historic city of Monterey, California, shares with St. Augustine, Florida, and Santa Fe, New Mexico, the distinction of possessing relics of a culture extending back to the Spanish régime in America. Here an enlightened community, assisted by the Carnegie Institution and the California State Park Commission, has invoked the aid of experts to study its past, and plan for the growth of the modern city in a way that will harmonize with perpetuation of those things that are significant in its history. Other historic monuments held by the state of California to commemorate important periods in its colorful past will benefit from the example of research and planning set at Monterey.

Closely related to the redwoods movement, there logically grew out of it a larger program for California involving preservation of many outstanding landscape resources—areas of mountain, forest, beach, and desert possessing uniqueness and distinction, not included in California's federal reservations or its outstanding national parks. From this developed a comprehensive statewide system of seventy state parks, with scenic, recreational, and historic values; the redwood parks taking their relative place in that varied pageant of beauty and interest which makes up California. This was one contribution of the League, and the first clear statement of its necessity was made by Dr. Merriam.

He never forgot his calling as a teacher. At the time when with Madison Grant and Henry Fairfield Osborn he formed the League and outlined its tasks, he was professor of palæontology at the University of California, with growing recognition in research, in teaching, and in administration. As a scientist he knew the importance of learning all the available facts before determining upon a course of action. As a teacher expounding to large classes a subject sometimes looked upon as abstruse, he realized the need of clear statement in terms of human values. As an administrator he understood the importance of bringing into coordinated action the efforts of many representing diverse points of view.

When called to the presidency of the Carnegie Institution of Washington, Dr. Merriam continued his participation in the Save-the-Redwoods movement, maintaining touch with it in the midst of a widening circle of affairs, and each summer (except one when he was abroad studying other nations' attempts to preserve the native landscape) returning to California on Institution business, and giving much of his time and thought to problems of the League.

He used to be a great fisherman, but he gave this up to spend his vacation time in saving the Redwoods.

It came to be felt that Dr. Merriam's annual appearance at the meeting of the Council of the Save-the-Redwoods League was necessary to retain the full measure of idealism in the efforts of those who had the good fortune to be associated with him in the cause. For another of his contributions was the establishment of a genuine and abiding enthusiasm for this movement among many associates—men of large business affairs, educators, scientists, foresters, landscape architects, public administrators, writers, artists, civic-minded women—an assemblage whom he once described as the outstanding group devoted to conservation in America.

Idealism. Stating a lofty purpose clearly and well. Striving for a goal of such magnitude as to grip the imagination. Taking the broad view, considering all interests involved. Not to compromise for temporary advantage. Never to take a defeatist attitude. These are some of the elements that entered into Dr. Merriam's leadership in the Save-the-Redwoods movement. The large financial assistance from men like John D. Rockefeller, Jr., and Edward S. Harkness, the generous appropriations by the state of California, the contributions by many individuals and organizations, and the sustained support by the several thousand members of the League, all followed a high conception and a clear statement of the ideals involved.

Nowhere is the significance of the movement better set forth than in his words at the dedication of the Franklin K. Lane Memorial Grove, in 1924:

> The movement for saving of the Redwoods, with which the dedication of this memorial is linked, has received such support as has rarely been given to any program of this nature. The purposes involved in the plan have extended themselves from the securing and protecting of limited tracts of forest to consideration of questions which concern the continuing growth of timber for future generations. They have touched even the most fundamental consideration of little known technical and scientific problems which must be settled before fullest realization of commercial values in reforestation and in forest protection can be attained.
>
> But with such vision as has been shown by those supporting the movement, and with the wholehearted support which has been furnished by every interest touching the work, we must realize that attainment of the objectives is only beginning. We are just now coming to understand the needs of state and nation

> which should be met by setting aside of forests for future recreational, scientific and æsthetic purposes; the needs of the lumberman for study of reforestation open an attractive and important field for future effort; and requirements of the community for a knowledge of ways in which the products of trees may be most useful furnish wide opportunity for application of the highest knowledge and skill.
>
> The success of the program, aimed to secure wisest use of this great asset of the country, depends ultimately upon interest and support of those who recognize our responsibilities to the future as well as to the present—who see that in development of the people of which we are a part there is now, as in every other age, much that should be done today and not tomorrow—much that if not done today shuts out forever great possibilities of further progress.
>
> We may not forget that in a true democracy the duty of citizenship extends beyond relations between members of a community at any given time. It involves a definite accountability to the past out of which we grew for all that the Creator and our ancestry have prepared for us in materials and in the power to use them. It includes the obligations to those arising from us and building on what will be their heritage.
>
> It is both the privilege and the duty of each generation to experience that joy which only the living of its own natural life can give. But no such pleasure is complete if based on the assumption that any generation lives for itself alone. A wide reach of opportunity for accomplishment is open to those who see that, small though their part may be, they have a place in that progressive movement of the world in which present links itself to past and future; who with an attitude of humility toward the power which moves the universe experience the joy of knowing that the heritage which they pass on is not less than what they have received.[3]

The human implications in an effort to hold intact and pass on to the future great values held in trust by this generation have ever been uppermost in Dr. Merriam's mind, and with the years have broadened both for him and for his associates. The scope of this conception was voiced by him while standing in the great forest cathedral of Bull Creek Flat at the time of its dedication as a public reserve in 1931; when at the same time the "Founders' Tree," world's tallest known standing tree, a redwood 364 feet high, was named in honor of the three men who established the League:

> The work of preservation is a result of effort by many individuals and organizations of this state and of the nation. There has never been a movement with greater appeal and finer support than that directed toward saving of the redwoods.

[3] John C. Merriam, "Dedication of the Franklin K. Lane Memorial Grove." Save-the-Redwoods League, 1924.

> The program included a wide variety of elements, ranging from uses which concern economic value up to the highest æsthetic and artistic aspects of appreciation.
>
> Although purely recreational value, in the sense of exercise and outdoor life, has had an important part in the movement to save these areas, this has always been subordinate to the idea that the redwoods are objects of exceptional beauty and primitive grandeur. It is the uses which satisfy the spirit of man rather than those which concern solely health of his body that have been dominant.
>
> The elusive element of natural beauty is that phase in the redwood program which to many seems of greatest importance. There is, however, another aspect which is universally a determining influence in the effort to save these groves; this is their primitive character. The fact that this is an unmodified result of creative activity, expressing development of life in one of its most splendid forms, has had great significance in work leading to preservation of the Bull Creek–Dyerville Forest.
>
> Beauty of nature is attractive, whether its development be guided artificially or is the direct result of growth without human influence. But the reaching upward of creative power and beauty out of a remote past gives to this forest the crowning influence of an otherwise magnificent natural spectacle. Our preservation of these groves we recognize as not merely the saving of a natural feature such as only centuries upon centuries could produce. We see it as a continuation of one of the most splendid aspects of creation in the world of living things.[4]

This sense of life as it has streamed through the ages has been at the bottom of Dr. Merriam's thinking throughout. He once remarked that his interest in the redwoods grew out of his early studies of the petrified redwood forest near Santa Rosa, California, where, turned to stone, there lie the trunks of ancient redwoods underneath a mass upon which grow the redwoods of today. This concept of time, this recognition of the "living link in an epic of history" represented by the remnant of widespread forests of millions of years ago, now making a last stand upon the northern coast of California, with all its meaning shines forth from the concluding passages of his chapter on the redwoods in his book "The Living Past":

> Among these deeper shades, in turning toward the world outside, you see the narrowing vistas ending not in shadow but in light. A flood of radiance sweeps against the vault above, penetrating here and there to give its living touch.
>
> Looking through crowded tree-tops, there is a glory in the forest that otherwise might remain unknown. The summits lift themselves to heights at which the voice of the wind in their branches is hushed by distance, and the delicacy of the

---

[4] John C. Merriam, Tallest Tree Dedicated, *American Forests*, November 1931.

topmost fronds seems refined to lace-like texture. Moving gently, the foliage touches across the openings, producing continuous variation of the light that streams or filters in. Beyond is the infinitely changing sky, a glowing sapphire through interlocking branches at midday; or with indescribably delicate tints if one sees it when level rays of morning or evening sun sweep over.

Standing in this field of shadow, among living relics of distant ages, in looking out you seem to turn from the clear story of a moving past to see the future rising from it through the miracle of never-failing light—the light that in unnumbered eons has poured down to mingle with the clouds of verdure, and build itself into the unfolding life and beauty of the forest.[5]

[5] *Idem*, "The Living Past," pp. 69–70. New York, Scribner, 1930.

JOHN CAMPBELL MERRIAM AS SCIENTIST AND PHILOSOPHER

CHESTER STOCK

California Institute of Technology
Research Associate, The Carnegie Institution of Washington

THE development of a man's philosophy when seen in the light of his past labor and experience frequently irradiates the process of thought itself and defines more clearly the contribution that is made in a search for the ultimate values of life. When one examines the writings of John Campbell Merriam the conclusion is inevitable that the richness and æsthetic qualities of his later contributions are the products of research and teaching in which human values have ever maintained an important place. To those who believe that a devotion to science and to its rigorous demands in the realm of thought presages a satisfactory approach to philosophy, it is stimulating and enlightening to trace the intellectual development of a great contributor to science.

Merriam's interest in the historical aspects of the geological sciences was evidently aroused at an early age, but the decision to make geology his chosen field for study, contemplation, and teaching formulated itself more definitely on completion of his undergraduate work at Lenox College, Iowa. His decision to come to the University of California in the late 'eighties and early 'nineties to study under Joseph LeConte emphasizes his growing desire and purpose, for Professor LeConte had become well known through his teachings and books as a scholar steeped in the lore of earth history. During his first period of residence at the University Merriam not only devoted himself to geological studies and served as assistant in mineralogy, but his range of interests included also the field of natural history. Under Professor E. L. Greene he was to acquire a foundation and sustaining interest in botany, which, as we shall see, became of considerable value and assistance to him in his later career as scientist and philosopher. Like many students of that day, Merriam took his doctorate abroad. At the University of Munich, under Professor Karl A. von Zittel, he selected palæontology as his major field of inquiry and wrote his dissertation "Ueber die Pythonomorphen der Kansas-Kreide," receiving his degree in 1893. Thus, a study of Cretaceous mosasaurs became a fortunate choice, for it permitted Merriam to obtain an acquaintance with the morphologic details of living and fossil reptiles, acquiring thereby a fund of information upon which he was to draw later in his study of the Triassic Ichthyosauria and Thalattosauria.

Although his doctorate was in vertebrate palæontology, Merriam had by this time acquired a very broad training in the geological sciences. On his return to the University of California in 1894 he was for some years occupied with both the invertebrate and vertebrate branches of palæontology. Between 1896 and 1899 he published six papers describing Tertiary molluscan faunas from Vancouver Island and discussing the geologic relations of the Martinez Group of California. His papers on the Tertiary echinoids of California remain perhaps his most significant contribution to invertebrate palæontology, for they laid the foundation on which important phylogenetic and stratigraphic studies were based by later investigators. Merriam was among the first to appreciate the significance of evolutionary changes in Tertiary echinoids, and especially their utilization in the recognition of horizon markers of value in an age determination and correlation of deposits in which these fossils are found.

In the course of his researches Merriam was frequently called upon to make identifications of fossil organisms encountered in Tertiary deposits exposed in the vicinity of San Francisco Bay. This led to an active cooperation with Professor A. C. Lawson in the study of the fossiliferous formations of the Concord quadrangle and in the unraveling of the geological history of that area. His first description of a fossil mammal from the California Tertiary was a result of this geological study. Prior to this description, he had already noted in print the occurrence of ichthyosaur remains in northern California, brought to his attention by the late James Perrin Smith. It was not until six years later, however, that he began to make available the results of his studies on this group of marine reptiles, and thence followed a series of papers which culminated in his great memoir on the Triassic Ichthyosauria, published in 1908.

This monograph treats not only of the geologic and geographic occurrences as well as of the morphological characters and taxonomy of the several types of Triassic ichthyosaurs known from America, but it also discusses succinctly the relationships of the Triassic types of western North America and the significance of these forms with regard to fundamental problems presented by higher vertebrates evolving under a marine environment. The development of the Ichthyosauria from antecedent land reptiles is generally accepted, but in the discussion of some factors involved in the evolution of the ichthyosaurs Merriam expresses clearly his palæontological point of view with regard to variation and environment. Thus, to quote from Merriam:

> The plasticity, or the greatest variation, in ichthyosaurs has appeared in those portions of the skeletal structure which stood in the most immediate relation to their environment, and could be most easily modified by changes in habits, food, or any variation in the elements of their surroundings. The most variable portions of the skeleton are probably the paddles and the caudal fin. In many of the principal characters of the cranial region of the skull a most remarkable conservatism is shown; only in the form of the rostrum and in the nature of the orbit do we find variation comparable to that of the limbs. The significance of the location of the centers of greatest variation seems to be that the variation is to a great extent determined by the peculiar relation of the animal to its environment, and in those centers where no distinctly important relation of this nature exists variation has been insignificant.

And again:

> The extent of the changes required in the accommodation of all ichthyosaurians, whenever and wherever they have existed, to those simpler features of aquatic life which are the same in all periods and in all places, may be responsible in a large measure for the generally similar trend of evolution.
>
> In this connection we naturally enquire why it was necessary for the Ichthyosauria to leave their original environment, for surroundings in which an entirely different equipment was necessary. The only explanation which offers itself suggests a possible change of food habits, through the discovery that certain of the thriving cephalopod groups of the early Mesozoic furnished an abundant and easily obtained food supply. Even in the most favorable light, however, the problem of definite variation in the Ichthyosauria in some of its aspects seems to invite a resort to orthogenesis; or to the view that there are factors influencing variation which are not as yet understood, and that these tendencies to similar variation may express themselves in related groups for long periods, and in regions widely separated geographically.

In passing it would be well to emphasize the difficulties inherent in this research on the ichthyosaurs. Only those who have had experience with the type of preservation of fossil remains encountered in the Triassic limestones of northern California can fully appreciate the tedious and painstaking application required in the identification of skeletal parts whose outlines are often well-nigh obliterated by long entombment of the fossil in the rock. Probably no other single palæontological contribution by Merriam is as formidable and as valuable in its documented record of an extinct group of organisms as this work on Triassic ichthyosaurs. The scope of the memoir makes it indispensable to future students of this great group of marine reptiles.

Contact with the intriguing problem of aquatic adaptation among marine reptiles and knowledge of the presence of fossil whales in the marine Tertiary of the Pacific Coast led Merriam to recognize the great importance of extended research on the extinct Cetacea. A program of investigation in this fertile field was at one time projected by Merriam, but pressure of other matters never permitted him actually to engage in this study. In the capable hands of one of his students, Remington Kellogg, this research is yielding results of outstanding scientific value.

With the closing years of the nineteenth century we see the initiation of a field of inquiry by Merriam destined to become one of his major contributions to science. The John Day Basin, lying east of the Cascades in north central Oregon, had yielded one of the more important faunal horizons in the American Tertiary sequence when Merriam and his expeditions explored this region in 1899 and 1900. The rich vertebrate life recorded particularly in the John Day beds had been described by the eminent palæontologists Cope, Marsh, and Wortman, but in no previous work had the geologic relationships of the several fossiliferous horizons been adequately established. This was accomplished by Merriam in his paper on the geology of the John Day Basin, a contribution that has furnished the basis for all further geologic and palæontologic inquiry in this section of the northwest. Not only was the geologic history and the sequence of faunas established by Merriam in this paper, but his observations on the origin and deposition of the John Day sediments were among the earliest to cast serious doubt on the prevailing views then held as to the lacustrine origin of many of the Tertiary deposits of the West. This enterprise was the first of many similar research projects that involved both field and laboratory investigations carried on by Merriam and in later years by his students under his guidance. Large portions of the Great Basin province of western North America were subjected to palæontological and geological inquiry with the result that many new Tertiary and Pleistocene vertebrate faunas were discovered.

In these studies Merriam frequently stressed the necessity of reading the geological story and the importance of determining the relation of a particular fragment of earth history to the account which might be rendered for a larger area. Thus, the investigations which were conducted by him in specific regions, as, for example, in the John Day Basin of Oregon, in the Thousand Creek and Virgin Valley basins of northwestern Nevada, or in the Mojave Desert area of southeastern California, to cite only a few, contributed important results to geology and palæontology. But as time and opportunity permitted these individual records of earth history were to be placed by Merriam and his associates in their proper time relations so that the broader correlations and generalizations might be realized in a recounting of the geological history for this great western region of the North American continent. We see here an early development of Merriam's views, expressed in later writings, wherein are unfolded the historical approach to problems, his constant desire to view the parts with regard to the whole, and his deep appreciation of time and movement in all change.

Merriam's studies in vertebrate palæontology were never dominated by a systematist's point of view. While he thoroughly appreciated and frequently availed himself of facts drawn from vertebrate zoology in an attempt to understand the biotic and physical environments under which fossil faunas lived, he was by no means content to describe fossil animals for the sake of recording only new species and genera. As a matter of fact a survey of his contributions to mammalian palæontology leaves the impression that he often conceded this privilege to others. His description of such forms was always undertaken with a view to making available data of value in geologic correlation, geographic distribution, or discussions of phylogeny. In his work on Tertiary mammals he seems seldom to have concerned himself with the central problems of evolution that have been of interest to palæozoologists. However, the problems with which he was concerned were in large measure geologic and the fragmentary character of the palæontological material afforded little basis for a discussion of such questions.

Yet his contributions to what might be designated the broader biological aspects of palæontology possess considerable scope and significance. Merriam's discussion of the relationship of fossil forms often displays the operation of a keenly analytical mind intent upon weighing facts derived from fragmentary evidence. Alternate possibilities are frequently in the balance and final judgment is often deferred until more and better materials become available. On occasion the presentation of evidence in the consideration of a problem reminds one of a legal brief. Relatively few of Merriam's papers contain restorations of the extinct animals he describes, illustrating his caution in pushing fact to the point of conjecture. Seldom, if ever, do we encounter a tendency to extend the palæontological generalizations beyond the realm of observational fact into that of speculation. When, for example, the curious antelopine genera *Ilingoceros* and *Sphenophalos* were discovered in the Thousand Creek Pliocene of northwestern Nevada, with their presumed important zoo-geographic implications, Merriam's description of these forms included a statement of their possible relationship to the Old World strepsicerine or twisted-horned antelopes. Previous work had suggested the occurrence of a form with these Eurasiatic affinities elsewhere in the Pliocene in America, but Merriam carefully refrained from subscribing to a view other than that both in the Old World and in North America there were present during the later Tertiary antelopine forms with twisted horns. Subsequently, when he explored the problem further, he was able to demonstrate the relationship of one of these genera (*Sphenophalos*) to the antilocaprid stock. Merriam's views concerning the twisted-horned antelopes involved a consideration of possibilities. To quote from him:

> In the present state of our knowledge there seem three hypotheses open to account for the presence of the twisted-horned antelopes in the Thousand Creek fauna: (1) They are typical Old World tragelaphines which came into America in late Miocene or early Pliocene time and developed long-crowned molar teeth. (2) They are tragelaphine forms which originated in America from Merycodus-like ancestors at some time during the Miocene, and soon migrated to the Old World, leaving only few descendants here as late as the Thousand Creek epoch.

> (3) They are a peculiar twisted-horned division of the Antilocapridae originating in America and possibly limited to this continent.

He concludes:

> It would seem to the writer that with the evidence available we are not in a position to determine the affinities of the American twisted-horned antelopes with certainty. So far as can be determined they appear to be near the Antilocapridae, but they are evidently generically distinct from *Sphenophalos*. If, as seems probable, the only type of dentition known from the Thousand Creek Beds really represents this group along with *Sphenophalos*, these forms are probably derived with the Antilocapridae from some type like *Merycodus*, and are not closely related to the Old World strepsicerine forms. If teeth like those obtained by Matthew and Cook at Snake Creek belong to this group it may represent an immigration of typical Old World forms or might be derived from a Palaeomeryx-like American form.
>
> With the available information it is probably desirable to refer *Ilingoceros* tentatively to a distinct family, the Ilingoceridae, and to include *Sphenophalos* in the Antilocapridae.

The occurrence of fossil mammalian assemblages in deposits intercalated in the sequence of marine sedimentary formations of the Coast Ranges of California furnished again a marvelous opportunity for the type of palæontological and geological investigation Merriam was particularly interested in. Furthermore, the problems of correlation presented here were of the utmost geologic interest since the position of these deposits offered for the first time an adequate basis for comparing the time relationships of the North American continental record of the later Tertiary with the marginal marine record of the Pacific Coast. Merriam's early broad training made him especially well fitted to consider this problem, which involved questions of vertebrate and invertebrate palæontology as well as of stratigraphy. His memoir on the "Tertiary Vertebrate Faunas of the North Coalinga Region of California," published in the *Transactions of the American Philosophical Society* in 1915, is a pivotal contribution wherein we note again his careful deductions from observed facts and his attempt to place this local history in its broader relationships.

With his knowledge of botany and as a result of his correlation studies in later years, Merriam was the first to recognize the significance of work in western palæobotany. Largely through his own efforts a research program in Cenozoic palæobotany was undertaken by Dr. Ralph W. Chaney, and the contributions which have come from Dr. Chaney and his associates have clearly shown the wisdom of developing palæontological inquiry in this direction.

The early search for ichthyosaurs in the Triassic limestones of northern California contributed indirectly to a wholly different type of palæontological inquiry. Recalling that Cope had obtained some Pleistocene bear material from a limestone cave in this region during the 'seventies, Merriam asked his assistant E. L. Furlong to explore the area for cavern occurrences. With the aid of an Indian guide, Furlong rediscovered Potter Creek Cave, and thus was initiated the program of cave explorations conducted under Merriam's direction for a number of years. Before completion of the excavations at Potter Creek Cave, Furlong had uncovered another occurrence, Samwel Cave, and an account of the exploration of this site was given some years later by Merriam in his charming story, "The Cave of the Magic Pool." In the progress of this work the question of man's antiquity arose, since in these cave deposits human remains and cultural materials were encountered, sometimes in close proximity to skulls and skeletal parts of animals that are generally looked upon as characteristic of the Ice Age.

Soon the program was expanded to include additional sites in other parts of the state. W. J. Sinclair, who took his doctorate under Merriam, completed the exploration of Potter Creek Cave and described the Pleistocene fauna from this and from other cavern deposits. Sinclair likewise concerned himself with the occurrences of early man in cave deposits and in the auriferous gravels of the Sierra Nevada. He later became professor of geology at Princeton University. In the region of San Francisco Bay the ancient shell mounds bordering the bay, in some instances giving evidence of accumulation prior to the last period of subsidence of that area, were investigated by Max Uhle and N. C. Nelson under the direction of Merriam.

The study of early man became for Merriam an ever enlarging sphere of interest. He recognized at an early stage the appeal which the subject has to those not intimately acquainted with the facts of palæontology and geology. In a popular article published in 1910 and entitled "The True Story of the Calaveras Skull" he discussed one of the early finds of supposedly ancient man recorded by J. D. Whitney from the Bret Harte country of the Sierra Nevada. This skull had attracted considerable attention because of assertions of its great antiquity by Whitney and others. The specimen itself gave definite evidence, however, of having come from some other deposit than that in which it was supposedly found. The probability that the Calaveras skull was a hoax, to confound the erudite if not wise professor who served as the second state geologist of California, may have been responsible, as Merriam notes in a later paper, for lack of credence in other occurrences of early man in America.

As a continuation of his interest in the problem of early man we note his statement published in 1910 on the relation of palæontology to the history of man, with particular reference to the American problem, presented in a symposium before the Society of Vertebrate Paleontologists. Here he called attention to the important part which palæontologists can take in the search for man's early records. His statement that "the results which must come are of great importance in the large problem of man's relation to nature" has been abundantly demonstrated by later research. It is not surprising to find, therefore, that when Merriam was asked to give the lectures in the William Ellery Hale Foundation before the National Academy of Sciences in 1918, he chose for the subject of his addresses: "The Beginnings of Human History Read from the Geological Record: The Emergence of Man." Here he reviews the palæontological evidence of man's origin and development in the changing environments of the geologic past and points to the need of appreciation of this background for proper understanding of the relationships of peoples in a modern world. He concludes with the statement:

> Man of the present day may read his story back to that early stage in which he first sees himself distinguished from the beast. He sees the beast made to a man-like beast and then a man. Perhaps to you the student of this ancient life has seemed to look upon a passing scene which might well have been left unknown —and yet to those who read what he who runs may see, the present world is brighter for the view—the future built upon the upward striving of the past must see the best there is in life at length prevail.

In more recent years Merriam reviewed the status of our knowledge concerning man's antiquity in America for the Sixteenth International Geological Congress and fostered a widespread attack on the problem through facilities offered by the Carnegie Institution in cooperation with other educational agencies. In this, as in other inquiries, Merriam conceives greatest effectiveness in reaching conclusions to be derived by interplay of several minds whose approach to the problem is through individual experience.

A large share of Merriam's time in the study of the Quaternary and its life was devoted to an exploration of the famous asphalt deposits of Rancho La Brea and to a description of the unusual treasures of prehistoric life recorded in these beds. Although this occurrence was known and had been mentioned in the literature as far back as the 'seventies, interest in the accumulation lay dormant until Merriam aroused widespread attention by describing the many unusual features which characterize these deposits and their entombed life record. The program of study and investigation which he instituted continues with unabated vigor. Its value when measured in terms of research and of scholarly productiveness by himself and by others is manifestly of a high order. In addition, Rancho La Brea is provocative to the imagination and has great public appeal.

A desire to share the facts derived from his special studies with a larger group than that comprised by the votaries of his own science is revealed by his essays for the lay reader. In addition to those already mentioned are others that also concern projects with which he has had intimate acquaintance, as for example, the article on "The John Day Fossil Beds" in *Harper's Magazine*, a description of the Rancho La Brea occurrence published in *Sunset* magazine and *Harper's Weekly*, and the several papers published in *Popular Science Monthly*. In his "Significant Features in the History of Life on the Pacific Coast," Merriam briefly outlines the more important stages in the history of plants and animals found in the fossil record of the Pacific Coast. Showing maturity of thought and deeper reflection are the later articles that have appeared in *Scribner's Magazine* and been reprinted in "The Living Past."

The urge to write a text which comes to every energetic teacher in a rapidly advancing field of inquiry was early thrust aside by Merriam. In its place came a synopsis of his lectures in palæontology which presents the subject matter in concise form and treats this with full cognizance of the fact that palæontology is closely affiliated with both the geological and the biological sciences. The synoptic treatment emphasizes Merriam's desire to have the student acquire knowledge not wholly by reading but by actual experience with the things as they are found in nature. And to Merriam experience implies contact with the outer world through as many sensory channels as one can marshal—certainly not that of sight alone.

It is not surprising, therefore, to find Merriam strongly advocating field as well as laboratory study and more particularly research. His energetic prosecution of the latter attracted a number of graduate students, especially from 1910 to the period of the World War, when he was in the full stride of his major scientific interest. His views concerning teaching and research are perhaps best expressed in his article on "The Function of Educational Institutions in Development of Research." Here he states:

> . . . the function of teaching in an educational institution does not concern alone the retailing of facts already assembled: it must include that kind of understanding of the subject which will prepare the student for his task as a leader in the future. To become such a leader that student must look beyond our present knowledge and experience with the expectation of accomplishing things which have never before been done. No good instructor can avoid recognizing this need of his students. No teacher who sees this requirement can fail to make a serious effort to determine the direction of advance in constructive use of his subject, if for nothing more than to point out to students the trend of the path and the preparation necessary for those by whom it will be extended to new fields of usefulness. It is hardly possible for the instructor to obtain a clear view of future development in his subject without intimate personal relation to the most advanced work in progress.

While he recognizes the value of the specialist in the acquisition of new knowledge, too much specialization often defeats its own end. The frontier posts in the advance of science are like salients developed along a battle front. From time to time these need to be consolidated, and only by an assimilation of the great and far-reaching discoveries by the masses, out of whom the great discoverers come, can we hope for an emancipation from wrong thinking and ignorance.

Merriam stresses on numerous occasions the need for a historical approach to our problems. As he remarks: "I am suggesting that the deepest view of history is desirable for purposes of most fundamental decisions; that, no matter how far back this vision leads us, if it continues to add to knowledge of what we are by showing us how we came to be, it is desirable and should be secured." He feels that the student of the earth sciences is in better position than anyone else to see the true course of history through the ages. Not only that, but the student comes to recognize the two great mobile elements in it, namely the earth itself and the stream of life which flows over it in time. Thus the earth sciences form, indeed, the background of history.

It is quite clear that Merriam's contributions to philosophic thought are the direct outgrowth of his work in science. They are the outcome of research and teaching in which human values have helped in the selection of those materials vital to teaching. He brings also a fundamental training in religion acquired in his youth, which with the passing of the years knows neither creed nor dogma. There is for him no conflict between religion and science, but both are essential to his philosophy of life. A long contact with the historical record, both physical and organic, convinces him that growth or evolution in the world is a continuing process. In this light, man and the world about him are the result of a long series of changes through the geological ages in which the present is merely one stage superimposed upon innumerable antecedent stages of the past. To him, the immeasurably long road over which this growth has come is in itself a certain assurance that evolution will not cease now but will continue into the future. The opposite view is as inconceivable as would be a concept that the earth, which has shown great change in the past, should forego all change in the future. And Merriam holds no brief for a world that is running down at the present time.

Unless some extraordinary forces appear, such as we have never been cognizant of before in the history of the world, it is inconceivable that the laws which lie behind the changes in the past will not also be operative in the future.

As Merriam passes in kaleidoscopic review this picture of the past he is particularly impressed by the reality of the story. It matters not that the evidence at times is fragmentary. The impress of feet in stone left ages ago, a leaf or a bone from strata formed in the geologic past mean to him the presence of once living organisms. More than that, they reveal an omnipotent creative force that has carried what we call life to higher and higher stages with the passing of geologic time. When, as sometimes happens, this story can be read at a great natural spectacle or shrine—at the brink of the Grand Canyon of the Colorado, in a redwood grove, or along the John Day River in eastern Oregon—it comes as revelation and with it is borne unto the mind the true meaning of time. And since in the main all growth or evolution has been forward and upward, are we to doubt that with the coming of the future there can be further advance? Long ago, the poet Bryant asked: "Has nature, in her calm, majestic march, faltered with age at last?" In his penetrating essay, "Are the Days of Creation Ended?" Merriam says:

> So, as we read the history of life and of man, and appreciate the dangers of our advance in learning and in civilization, we may realize that we are taking for our use that dangerous fruit from the tree of knowledge, and may question whether we can bear the consequence of coming to view the world as gods. At such a time it is proper that we turn in reverential inquiry to view the record of the past through which creation has revealed itself. There we find a history open for our information. It tells us of forward movement of the world from which we have grown, and in which we are a part. What greater gift could we receive than to learn through the evidence there presented that the conditions which have governed the creative process are essentially those which operate to-day and should be expected to continue; that whatever power has existed behind nature in the past expresses itself to-day, keeping the world a place that may be ever new, and justifying faith that progress will extend itself into the future?

This story of the past holds for man of today a special significance, since it is he who completes the link between past and future. Man faces thereby a sacred trust because of his dominant position. He must needs be called upon to judge those values of peculiar import to his place in the continuity of life, for he cannot live by self alone. Science may help, but in the end one of its greatest contributions should be in the determination or selection of proper human values. In conservation, in the maintenance of life, in the appreciation of nature, in keeping men free, we may look with confidence to science for help. As Merriam says:

> . . . even more important than those values which aid to bring riches and abolish poverty is the fact that science contributes toward betterment of life by placing before us the desirability of facing the realities, or the facts, and of taking the broader and longer view with reference to everything in life. Giving increased significance to truth, and depending more fully upon it, means not only a bettered and more nearly stable and safer world, but one in which the possibility of forward building or of accomplishment is enormously enhanced.

Our spiritual values should likewise never diminish, for as Merriam states:

> No one can escape the feeling that, as knowledge advances, the greater world we come to know in terms of space, and time, and power, becomes the object of increasing reverence. Our place of habitation and the period of our existence shrink to relatively small dimensions, but the glory of our part in the plan of the universe increases when we see creation renewed in each successive age, and the way open to us for that continuing growth or progress to which the life of each stage has made its contribution.

So I repeat what a poet and philosopher has written for us:

> Let knowledge grow from more to more,
> But more of reverence in us dwell;
> That mind and soul, according well,
> May make one music, as before.

And again:

> One of the greatest advances of all time was that expressed, ages ago, in the view that there is in the universe one power in many forms, or, in terms of religion, one God instead of many warring deities. It may be in order for mankind to make this discovery anew, or from time to time, when unity in views of the world and in belief seems threatened by erection of too many temples to deities of varying and perhaps inconsistent missions, in a world that, so far as nature is concerned, has operated as one system since time's beginning.

INDEX

Aboriginal American History, Section of, 728; *see also* Historical Research, Division of, *and* Maya Researches
Adams, L. H., vi
  Geophysical studies, 37–47
Adams, Walter S., vi
  Astronomical studies, 135–147
Administration Building of Carnegie Institution of Washington, 748, 750, 751
Advisory Committee on Seismology, work of, 3–33, 728
Animal Biology, Division of, 397, 728
  Studies in, *see* Embryology, Genetics, Nutrition
Astrometry, Department of Meridian, 749
Astronomy, magnitudes and colors of stars, 149–159
  Moon surface, 59–74
  Nebulæ, 91–102
  Photoelectric cell in, 75–89
  Planets, 115–134
  Solar and terrestrial relationships, 103–114
  Sunspots and stellar distances, 135–147
  See also Mount Wilson Observatory
Atmospheric Electricity, 183–204, 226–238
Atomic Physics, studies in, 161–181
Atomic Physics Observatory, 180–181
Avery, Amos G., *Datura* investigations, 315–351

Benedict, Francis G., Nutrition Laboratory researches, 251–258
Berkner, L. V., variations in terrestrial magnetism, 223–247
Bibliographies, 33–35, 203–204, 246–247, 271–273, 292–293, 312–314, 346–351, 367, 379–382, 387, 413, 510, 562–565, 583–584, 619–620, 648, 680–682, 694
Blakeslee, Albert F., *Datura* investigations, 315–351
Bliss, Robert Woods, vi
Blood Cells, investigations of, 369–382
Bunker, Frank F., vi
  Conduct and interpretation of research, 713–752
Buwalda, John P., geology of part of Pacific Northwest, 695–710

California, earthquake studies in, 3–33
  Saving the redwoods, 753–763
Carnegie, the, 186, 191, 192, 198–200, 207, 208
Carnegie, Andrew, 59, 714, 715, 726, 729
Carnegie Corporation of New York, 15, 738
Carnegie Institution, 3–7, 23, 26, 30, 34, 59–60, 74, 88, 103, 186, 225, 229, 260, 302, 315, 316, 400, 415, 433, 438, 439, 446, 450, 465, 536, 623, 624, 632, 650, 671, 673, 676, 685, 713, 716, 717, 724, 726, 727, 728, 729, 733, 735–737, 742–752, 759
Carpenter, Thorne M., studies on respiration, 425–433
Cetacea, *see* Whales
Chaney, Ralph W., ancient forests of Oregon, 631–648
Christianity, introduction into Yucatan, 585–620
Chromosomes, studies of, 275–293, 295–314, 315–351
Chromospheric Eruptions, *see* Sun
Chronologies, Maya and Christian, 493–510
Conservation of the California Redwoods, 753–763
Cosmic-Ray Investigations, Committee on Coordination of, 236, 728

Datura stramonium, studies of, 315–351
Day, Arthur L., studies in seismology, 3–35
Delano, Frederic A., vi
Delphinium, genetic researches on, 296–301
Demerec, M., research on the gene, 295–314
Density, geophysical investigations of, 39–41
Drosophila, studies of, 276–287, 298–312
Drury, Newton Bishop, conserving the California redwoods, 753–763

Dunham, Theodore, Jr., studies of the planets, 115–134

Early Man in America, 623–629, 772–774
Earth, *see* Astronomy *and* Magnetic Studies
Earthquakes, *see* Seismology
Electricity, atmospheric, 183–204, 226–238
Embryology, Department of, 275, 369, 383, 389, 398, 399, 433, 671, 728, 749
  Studies in, 275–292, 383–387, 389–396, 397–413
Equilibrium, in geophysical investigations, 44–45

Fleming, J. A., vi
  Studies of magnetism, 205–221
Flies, fruit, *see Drosophila*
  Fungus, *see Sciara*
Fruit flies, *see Drosophila*
Fungus flies, *see Sciara*

Gene, research on, 295–314
  See also Genetics
Genetics, Department of, 259, 260, 295, 297, 298, 315, 353, 355, 433, 567, 728, 749
  Studies in, 259–271, 275–292, 295–314, 315–351, 353–367, 567–584
Geophysical Investigations, pressure and volume in, 37–47
  Water in geological processes, 49–58
Geophysical Laboratory, 3, 12, 37, 49, 55–59, 162, 216, 219, 726, 749
Geological Processes, water in, 49–58
Geology, contributions of J. C. Merriam, 765–769
  Studies in, 631–648, 683–694, 695–710
Gilbert, W. M., vi
Gish, O. H., atmospheric electricity, 183–204
Grand Canyon, geological formations in, 688–693
Guatemala, researches in, *see* Maya Researches

Hafstad, L. R., studies of magnetism, 161–181
Hartman, Carl G., studies in embryology, 389–396
Heuser, Chester H., studies in embryology, 383–387
Heydenburg, N. P., studies of magnetism, 161–181
Historical Research, Division of, 465, 493, 533, 585, 728, 749
History, American, 483–492, 493–510, 511–532, 533–565, 585–620
  Of science, 465–481
Howard, E. B., studies of Early Man in America, 623–629
Huancayo Observatory, 26, 109, 188, 210, 225, 229
Hubble, Edwin, studies of nebulæ, 91–102

Idolatry in Ancient Yucatan, 585–620
Indians, *see* Maya Researches

John Day Region, geology of, 696–710

Kellogg, Remington, adaptation in whales, 649–682
Kidder, A. V., vi
  Maya and Christian Chronologies, 493–510

Landa, and idolatry in Yucatan, 585–620
Leland, Waldo G., substance of American history, 483–492
Leukemia, studies of, 353–367
Lewis, Margaret R., studies on white blood cells, 369–382
Lewis, Warren H., studies on white blood cells, 369–382
Literature Cited, *see* Bibliographies
Lunar Studies, *see* Moon.

Macaca mulatta, see Monkeys
MacDowell, E. C., studies of leukemia, 353–367
McKee, Edwin D., geological studies, 683–694
McNish, A. G., variations in terrestrial magnetism, 223–247
Magmas, water in, 49–52, 55–58
Magnetic Studies, 106–114, 161–181, 183–204, 205–221, 223–247
Maize, genetic researches on, 276, 295–296, 298
Man, early, 623–629, 772–774
Marine Biological Laboratory, 728, 749
Marine Biology, Department of, 726

Maya Researches, anthropological studies, 567–584
  Chronologies, 493–509
  Classes, 511–532
  Landa and idolatry, 585–620
  New Empire, 533–565
Meridian Astrometry, Department of, 749
Merriam, John Campbell, v, vi, 3, 5, 32, 59, 60, 74, 315, 316, 390, 465, 623, 624, 625, 632, 633, 650, 651, 672, 696, 697, 703, 706, 707, 725–730, 737–742, 751, 752, 753–763, 765–777
  And the redwoods, 753–763
  As scientist and philosopher, 765–778
  Quoted, 725–727, 729, 730, 737–742, 752, 756–757, 760–763
Metabolism, 253–257, 262, 264, 575
Metz, C. W., chromosome studies of *Sciara*, 275–293
Mice, study of tumors in, 353–367, 377–379
Monkeys, studies of embryology of, 383–387, 389–396, 397–413
Moon, Committee on Study of Surface Features of the, 59–74, 728
  Studies of, 59–74; *see also* Astronomy
Morey, George W., studies on water in geological processes, 49–58
Morley, Sylvanus G., Maya studies, 533–565
Mount Wilson Observatory, 7, 10, 18, 60, 69, 73, 76, 78, 81, 86, 89, 91–93, 98–99, 101, 103, 105, 108–109, 115, 120–122, 124–125, 127–130, 135, 137–139, 141, 142, 144, 146, 149, 154–159, 726, 749, 750

Nebulæ, studies of, 91–102; *see also* Astronomy
Nicholson, Seth B., solar and terrestrial relations, 103–114
Nutrition Laboratory, 399, 415, 425, 749
  Researches of the, 251–258, 728
Nutrition Studies, 251–258, 415–423
Oregon, ancient forests of, 631–648
  Geology of, 695–710

Palæobotany, studies in, 631–648
Palæontology, contributions of J. C. Merriam, 766–775
  Studies in, 623–629, 631–648, 649–682
Photoelectric Cells in Astronomy, 66–68, 75–89
Photosynthesis, 453–461, 716–724, 732–733,
Pigeons, changing organism in, 260–271
Planets, studies of, 115–134
Plant Biology, Division of, 435, 440, 676, 716, 723, 728, 749
  Studies in, 435–462, 716–724
Post-Columbian American History, Section of, 728; *see also* Historical Research, Division of
President's Reports, excerpts from, 725–727, 729, 730, 737–742, 752
Press, as distributor of knowledge, 745–748
Pressure, in geophysical investigations, 37–47
Publications, Office of, 713

Races in Yucatan, 511–532
Redfield, R., race and class in Yucatan, 511–532
Redwoods, conserving the California, 753–763
Reproductive Cycle in Pigeons, 266–267
Research, conduct and interpretation of, 713–752
Respiration, studies on, 425–433
Respiratory Exchange, effects of changes in breathing, 425–433
Riddle, Oscar, genetic researches, 259–273
Root, Elihu, 729
  Lectures, 731, 751
Roys, Ralph L., studies in Maya history, 585–620

Sarton, George, history of science, 465–481
Save-the-Redwoods League, 753–763
Scholes, France V., studies in Maya history, 585–620
Schools, as distributors of knowledge, 742–745
Sciara, chromosome studies, 275–293
Science, conduct and interpretation of research in, 713–752
  History of, 465–481
  Section of History of, 728; *see also* Historical Research
Seares, Frederick H., studies of stars, 149–159
Sedimentary Rocks, studies of, 683–694

Seismology, Advisory Committee on, 3–33, 728
  Studies in, 3–35
Sequoia, *see* Redwoods
Sherman, Henry C., nutrition studies, 415–423
Spectra, stellar, 136–147
Spoehr, H. A., vi
  Plant biology studies, 435–462, 716–724
  Quoted, 731
Stars, *see* Astronomy
Stebbins, Joel, photoelectric cells in astronomy, 75–89
Steggerda, Morris, Maya Indians of Yucatan, 567–584
Stock, Chester, tribute to John Campbell Merriam, 765–778
Streeter, George L., vi
  Studies in embryology, 397–413
Sun, and magnetic disturbances on the earth, 103–114, 229–246
  Sunspots, 103–106, 135–147
  See also Astronomy
Sunspots, 103–106, 135–147; *see also* Sun

Terrestrial Magnetism, Department of, 26, 28, 66, 109, 161–162, 165, 167–169, 177, 180, 183, 186–188, 205–208, 210, 211, 213, 216, 217, 219, 220, 223, 225, 226, 228, 229, 231, 236, 238, 241–245, 749
Thompson, J. E., Maya and Christian chronologies, 493–510
Tippler Pigeons, study of, 260–271
Tuve, M. A., studies of magnetism, 161–181

United States History, Section of, 728; *see also* Historical Research

Volcanic Disturbances, 49, 53–54, 631–632, 647; *see also* Seismology
Volume, in geophysical investigations, 37–47

Water, in geological processes, 49–58
Watheroo Observatory, 109, 188, 210, 225, 243, 244
Weed, Lewis, H., vi
Whales, adaptation of structure to function in, 649–682
Wright, F. E., vi
  Studies of moon, 59–74

Yucatan, research in, *see* Maya Researches

SMITHSONIAN INSTITUTION LIBRARIES
3 9088 00735 8302